

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:16:52 ; Search time 2286.42 Seconds  
(without alignments)  
49.243 Million cell updates/sec

Title: US-09-380-826A-2  
Perfect score: 22  
Sequence: 1 ttttgatcacaaagtattgata 22

Scoring table: IDENTITY\_NOC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_ph.\*
- 6: gb\_pl1.\*
- 7: gb\_pl2.\*
- 8: gb\_pr1.\*
- 9: gb\_pr2.\*
- 10: gb\_pr3.\*
- 11: gb\_ro.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: em\_fun.\*
- 15: em\_hum1.\*
- 16: em\_hum2.\*
- 17: em\_in.\*
- 18: em\_om.\*
- 19: em\_or.\*
- 20: em\_ov.\*
- 21: em\_pat.\*
- 22: em\_ph.\*
- 23: em\_pl.\*
- 24: em\_ro.\*
- 25: em\_sts.\*
- 26: em\_sy.\*
- 27: em\_un.\*
- 28: em\_v1.\*
- 29: gb\_ba3.\*
- 30: gb\_in1.\*
- 31: gb\_in2.\*
- 32: gb\_in3.\*
- 33: gb\_pl3.\*
- 34: gb\_pr4.\*
- 35: em\_ba1.\*
- 36: em\_ba2.\*
- 37: em\_htg1.\*
- 38: em\_htg2.\*
- 39: em\_htg3.\*
- 40: em\_htg4.\*
- 41: em\_htg5.\*
- 42: em\_htg6.\*
- 43: em\_htg7.\*

- 44: em\_htg8.\*
- 45: em\_htg9.\*
- 46: em\_htg10.\*
- 47: em\_hum3.\*
- 48: em\_hum4.\*
- 49: em\_hum5.\*
- 50: em\_hum6.\*
- 51: gb\_pr5.\*
- 52: gb\_pr6.\*
- 53: gb\_pr7.\*
- 54: gb\_htg1.\*
- 55: gb\_htg2.\*
- 56: gb\_htg3.\*
- 57: gb\_htg4.\*
- 58: gb\_htg5.\*
- 59: gb\_htg6.\*
- 60: gb\_htg7.\*
- 61: gb\_htg8.\*
- 62: gb\_htg9.\*
- 63: gb\_htg10.\*
- 64: gb\_htg11.\*
- 65: gb\_htg12.\*
- 66: gb\_htg13.\*
- 67: gb\_htg14.\*
- 68: gb\_htg15.\*
- 69: gb\_htg16.\*
- 70: gb\_htg17.\*
- 71: gb\_htg18.\*
- 72: gb\_htg19.\*
- 73: gb\_htg20.\*
- 74: gb\_htg21.\*
- 75: gb\_htg22.\*
- 76: gb\_htg23.\*
- 77: gb\_sts1.\*
- 78: gb\_sts2.\*
- 79: gb\_vil.\*
- 80: gb\_vil2.\*
- 81: gb\_pat1.\*
- 82: gb\_pat2.\*
- 83: em\_htg0.\*
- 84: gb\_htg24.\*
- 85: gb\_pr8.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	%	Query Match	Length	DB	ID	Description
1	22	100.0	1481	2	LFU60594			U60594 Leptospi
2	20.4	92.7	354	1	AB007014			AB007014 Spirochae
3	18.8	85.5	283	1	AF003953			AF003953 Leptospi
4	18.8	85.5	288	2	LIU94974			U94974 Leptospi
5	18.8	85.5	288	2	LIU94975			U94975 Leptospi
6	18.8	85.5	288	2	LIU94976			U94976 Leptospi
7	18.8	85.5	288	2	LIU94977			U94977 Leptospi
8	18.8	85.5	288	2	LIU94978			U94978 Leptospi
9	18.8	85.5	288	2	LIU94979			U94979 Leptospi
10	18.8	85.5	353	1	AB007012			AB007012 Spirochae
11	18.8	85.5	1012	7	ATHRNS2X			M98336 Arabidopsi
12	18.8	85.5	72590	6	AC003000			AC003000 Arabidops
13	18.8	85.5	109723	72	AF165146			AF165146 Homo sapi
14	18.8	85.5	200799	6	AF137379			AF137379 Nephrocel
15	18.8	85.5	200799	6	AF137379			AF137379 Nephrocel
c 16	18.4	83.6	1812	31	AF220067			AF220067 Drosophil
c 17	18.4	83.6	27365	58	AC015209			AC015209 Drosophil
c 18	18.4	83.6	185087	54	AC007475			AC007475 Drosophil
19	18.4	83.6	262731	30	AE003823			AE003823 Drosophil
20	17.8	80.9	1874	79	EHBP44NCP			L48441 Echinocloa
c 21	17.8	80.9	31214	33	SPCC830			AL109850 S.pombe c

Wed Mar 28 14:03:14 2001

AB009052 Arabidops  
AC026327 Homo sapi  
AC026320 Homo sapi  
AC055742 Homo sapi  
AC003114 Arabidops  
AB008265 Arabidops  
AC068144 Homo sapi  
AC021015 Homo sapi  
AC023155 Homo sapi  
AF175672 Unculture  
AF271231 Albugo ca  
AF018567 Unidentif  
D49367 Lithospermu  
Z73136 S.cerevisia  
U40832 Strongyloce  
U41016 Caenorhabdi  
U41545 Caenorhabdi  
AL049803 Arabidops  
AC083790 Homo sapi  
AC036233 Homo sapi  
AP002082 Homo sapi  
AL078469 Arabidops  
AL078470 Arabidops  
Z98052 Human DNA s

17.8 80.9 85992 6 AB009052  
17.8 80.9 91894 65 AC026327  
17.8 80.9 174766 65 AC026320  
17.8 80.9 270889 33 AC055742  
17.4 79.1 59261 33 T12M4  
17.4 79.1 81652 6 AB008265  
17.4 79.1 109512 68 AC068144  
17.4 79.1 149241 61 AC021015  
17.4 79.1 187847 62 AC023155  
17.2 78.2 462 1 AF175672  
17.2 78.2 835 7 AF271231  
17.2 78.2 1083 1 AF018567  
17.2 78.2 2213 33 LEP4CCOALB  
17.2 78.2 3608 33 SCYLL031C  
17.2 78.2 5596 32 SPUA0832  
17.2 78.2 34496 31 CEUR11G1  
17.2 78.2 38655 31 CEUC02F12  
17.2 78.2 49311 84 F26K10  
17.2 78.2 62478 71 AC083790  
17.2 78.2 67858 66 AC036233  
17.2 78.2 68040 76 AP002082  
17.2 78.2 97714 7 ATF25024  
17.2 78.2 104679 7 ATF19B15  
17.2 78.2 112467 52 HS0505B13

ALIGNMENTS

RESULT 1  
LFU60594 1481 bp DNA BCT 10-SEP-1998  
LOCUS Leptospira fainei 16S ribosomal RNA gene, partial sequence.  
AC0594  
U60594.1 GI:1408219  
Leptospira fainei.  
Leptospira fainei.  
Bacteria: Spirochaetales; Leptospiraceae; Leptospira.  
REFERENCE 1 (bases 1 to 1481)  
AUTHORS Perolat, P., Chappel, R.J., Adler, B., Baranton, G., Bulach, D.M.,  
Billinghurst, M.L., Letocart, M., Merien, F. and Serrano, M.S.  
Leptospira fainei sp. nov., isolated from pigs in Australia  
Int. J. Syst. Bacteriol. 48 Pt 3, 851-858 (1998)  
JOURNAL 98404550  
MEDLINE 2 (bases 1 to 1481)  
REFERENCE Adler, B., Chappel, R.J., Baranton, G., Bulach, D.M.,  
Billinghurst, M.L., Letocart, M., Merien, F., Serrano, M.S. and  
Perolat, P.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUN-1996) Microbiology, Monash University, Wellington  
RD., Clayton, VIC 3168, Australia  
FEATURES  
source Location/Qualifiers  
1..1481  
/organism="Leptospira fainei"  
/strain="Hurstbridge"  
/db\_xref="taxon:48782"  
1..>1481  
/product="16S ribosomal RNA"  
BASE COUNT 391 a 335 c 439 g 314 t 2 others  
ORIGIN  
Query Match 100.0%; Score 22; DB 2; Length 1481;  
Best Local Similarity 100.0%; Pred. No. 3.9;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 ttttgatcacaaagattgata 22  
|||||  
Db 155 TGTGGATCACAAAGATTGATA 176  
|||||

AB007014 354 bp DNA BCT 13-OCT-1997  
LOCUS Spirochaeta sp. 16S rRNA gene, partial sequence.  
DEFINITION AB007014  
AC007014  
AB007014.1 GI:2516255  
VERSION 16S ribosomal RNA.  
KEYWORDS Spirochaeta sp. (sub-species: Freshwater obligate oligotroph,  
strain: FO-95) DNA.  
SOURCE Spirochaeta sp.  
ORGANISM Spirochaetales; Spirochaetales; Spirochaetales; Spirochaeta.  
REFERENCE 1 (bases 1 to 354)  
AUTHORS Shin, M.-S.  
TITLE Direct Submission  
JOURNAL Submitted (05-SEP-1997) to the DDBJ/EMBL/GenBank databases. Mi-Sun  
Shin, Laboratory of Marine Molecular Microbiology, Faculty of  
Agriculture, Kyoto University; Oiwakecho, Kitashirakawa, Sakyo-ku,  
Kyoto, Kyoto 606-01, Japan (E-mail: mlsun@kais.kyoto-u.ac.jp,  
Tel: 075-753-6224, Fax: 075-753-6226)  
2 (sites)  
AUTHORS Shin, M., Yoshinaga, I., Uchida, A. and Ishida, Y.  
TITLE Phylogenetic analysis by 16S rRNA gene sequencing of Obligate  
Oligotrophs isolated from the northern basin of Lake Biwa  
(Mesotrophic Lake)  
JOURNAL Unpublished (1997)  
FEATURES  
source Location/Qualifiers  
1..354  
/organism="Spirochaeta sp."  
/strain="FO-95"  
/sub-species="Freshwater obligate oligotroph"  
/db\_xref="taxon:28185"  
1..>354  
/product="16S rRNA"  
BASE COUNT 91 a 79 c 111 g 73 t  
ORIGIN  
Query Match 92.7%; Score 20.4; DB 1; Length 354;  
Best Local Similarity 95.5%; Pred. No. 25;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 ttttgatcacaaagattgata 22  
|||||  
Db 133 TGTGGATCACAAAGATTGATA 154  
|||||

RESULT 3  
AF003953 283 bp DNA BCT 30-MAY-1998  
LOCUS Leptospira interrogans strain 48/95 16S ribosomal RNA gene, partial  
DEFINITION sequence.  
AC003953  
AF003953.1 GI:3169306  
VERSION  
KEYWORDS  
SOURCE Leptospira interrogans.  
ORGANISM Leptospira interrogans  
Bacteria: Spirochaetales; Leptospiraceae; Leptospira.  
REFERENCE 1 (bases 1 to 283)  
AUTHORS Woo, T.H.S., Patel, B.K.C., Cinco, M., Smythe, L.D., Symonds, M.,  
Norris, M. and Dohnt, M.  
TITLE Identification of Leptonema by real-time homogeneous assay of rapid  
cycle PCR product  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 283)  
AUTHORS Woo, T.H.S., Patel, B.K.C., Cinco, M., Smythe, L.D., Symonds, M.,  
Norris, M. and Dohnt, M.  
TITLE Direct Submission  
JOURNAL Submitted (14-MAY-1997) School of Science, Griffith University,  
Brisbane, QLD 4111, Australia  
FEATURES  
source Location/Qualifiers  
1..283  
/organism="Leptospira interrogans"  
/strain="48/95"  
/db\_xref="taxon:173"

```

REFERENCE          2 (bases 1 to 288)
AUTHORS            Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
                   Brenner,D.J. and Patel,B.K.C.
TITLE              Direct Submission
JOURNAL            Submitted (24-MAR-1997) School of Science, Griffith University,
                   Brisbane, QLD 4111, Australia
FEATURES           Location/Qualifiers
                    source          1..288
                                     /organism="Leptospira inadai"
                                     /strain="68/94"
                                     /db_xref="taxon:29506"
                                     <1..>288
rRNA               /product="16S ribosomal RNA"
BASE COUNT        77 a      61 c      88 g      62 t
ORIGIN             1..288
                  ttgttgatcacaaagtattgata 22
                  ||| |||| |||| | |||| ||||
Db 141 TATTGGATCAGGATTGTGATA 162

RESULT            6
LIU94976          LIU94976      288 bp     DNA      BCT       01-JAN-1998
LOCUS              Leptospira inadai 16S ribosomal RNA gene, partial sequence.
DEFINITION         U94976
ACCESSION          U94976
VERSION            U94976.1 GI:2735448
KEYWORDS            Leptospira inadai.
SOURCE             Leptospira inadai.
ORGANISM           Bacteria: Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE          1 (bases 1 to 288)
AUTHORS            Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
                   Brenner,D.J. and Patel,B.K.C.
TITLE              Identification of Leptospira inadai by continuously monitoring
                   fluorescence during rapid cycle PCR
JOURNAL            Unpublished
REFERENCE          2 (bases 1 to 288)
AUTHORS            Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
                   Brenner,D.J. and Patel,B.K.C.
TITLE              Direct Submission
JOURNAL            Submitted (24-MAR-1997) School of Science, Griffith University,
                   Brisbane, QLD 4111, Australia
FEATURES           Location/Qualifiers
                    source          1..288
                                     /organism="Leptospira inadai"
                                     /strain="346/95"
                                     /db_xref="taxon:29506"
                                     <1..>288
rRNA               /product="16S ribosomal RNA"
BASE COUNT        77 a      61 c      88 g      62 t
ORIGIN             1..288
                  ttgttgatcacaaagtattgata 22
                  ||| |||| |||| | |||| ||||
Db 141 TATTGGATCAGGATTGTGATA 162

Query Match      85.5% ; Score 18.8 ; DB 2 ; Length 288 ;
Best Local Similarity 90.9% ; Pred. No. 1.4e+02 ;
Matches 20 ; Conservative 0 ; Mismatches 2 ; Indels 0 ; Gaps 0 ;

Qy 1 ttgttgatcacaaagtattgata 22
    ||| |||| |||| | |||| ||||
Db 141 TATTGGATCAGGATTGTGATA 162

RESULT            7
LIU94977          LIU94977      288 bp     DNA      BCT       01-JAN-1998
LOCUS              Leptospira inadai 16S ribosomal RNA gene, partial sequence.
DEFINITION         U94977
ACCESSION          U94977
VERSION            U94977.1 GI:2735448
KEYWORDS            Leptospira inadai.
SOURCE             Leptospira inadai.
ORGANISM           Bacteria: Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE          1 (bases 1 to 288)
AUTHORS            Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
                   Brenner,D.J. and Patel,B.K.C.
TITLE              Identification of Leptospira inadai by continuously monitoring
                   fluorescence during rapid cycle PCR
JOURNAL            Unpublished
REFERENCE          2 (bases 1 to 288)
AUTHORS            Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
                   Brenner,D.J. and Patel,B.K.C.
TITLE              Direct Submission
JOURNAL            Submitted (24-MAR-1997) School of Science, Griffith University,
                   Brisbane, QLD 4111, Australia
FEATURES           Location/Qualifiers
                    source          1..288
                                     /organism="Leptospira inadai"
                                     /strain="346/95"
                                     /db_xref="taxon:29506"
                                     <1..>288
rRNA               /product="16S ribosomal RNA"
BASE COUNT        77 a      61 c      88 g      62 t
ORIGIN             1..288
                  ttgttgatcacaaagtattgata 22
                  ||| |||| |||| | |||| ||||
Db 141 TATTGGATCAGGATTGTGATA 162

Query Match      85.5% ; Score 18.8 ; DB 2 ; Length 288 ;
Best Local Similarity 90.9% ; Pred. No. 1.4e+02 ;
Matches 20 ; Conservative 0 ; Mismatches 2 ; Indels 0 ; Gaps 0 ;

Qy 1 ttgttgatcacaaagtattgata 22
    ||| |||| |||| | |||| ||||
Db 141 TATTGGATCAGGATTGTGATA 162

RESULT            7
LIU94977          LIU94977      288 bp     DNA      BCT       01-JAN-1998
LOCUS              Leptospira inadai 16S ribosomal RNA gene, partial sequence.
DEFINITION         U94977
ACCESSION          U94977
VERSION            U94977.1 GI:2735448
KEYWORDS            Leptospira inadai.
SOURCE             Leptospira inadai.
ORGANISM           Bacteria: Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE          1 (bases 1 to 288)
AUTHORS            Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
                   Brenner,D.J. and Patel,B.K.C.
TITLE              Identification of Leptospira inadai by continuously monitoring
                   fluorescence during rapid cycle PCR
JOURNAL            Unpublished
REFERENCE          2 (bases 1 to 288)
AUTHORS            Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
                   Brenner,D.J. and Patel,B.K.C.
TITLE              Direct Submission
JOURNAL            Submitted (24-MAR-1997) School of Science, Griffith University,
                   Brisbane, QLD 4111, Australia
FEATURES           Location/Qualifiers
                    source          1..288
                                     /organism="Leptospira inadai"
                                     /strain="346/95"
                                     /db_xref="taxon:29506"
                                     <1..>288
rRNA               /product="16S ribosomal RNA"
BASE COUNT        77 a      61 c      88 g      62 t
ORIGIN             1..288
                  ttgttgatcacaaagtattgata 22
                  ||| |||| |||| | |||| ||||
Db 141 TATTGGATCAGGATTGTGATA 162

```

Wed Mar 28 14:03:14 2001

us-09-380-826a-2.rge

U94977.1 GI:2735449

VERSION

KEYWORDS

SOURCE Leptospira inadai.

ORGANISM

Bacteria; Spirochaetales; Leptosiraceae; Leptospira.

REFERENCE

1 (bases 1 to 288)

AUTHORS

Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M., Brenner, D.J. and Patel, B.K.C.

TITLE

Identification of Leptospira inadai by continuously monitoring

JOURNAL

fluorescence during rapid cycle PCR

REFERENCE

2 (bases 1 to 288)

AUTHORS

Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M., Brenner, D.J. and Patel, B.K.C.

TITLE

Direct Submission

JOURNAL

Submitted (24-MAR-1997) School of Science, Griffith University, Brisbane, QLD 4111, Australia

FEATURES

Location/Qualifiers

source

1. .288

/organism="Leptospira inadai"

/strain="268/95"

/db\_xref="taxon:29506"

&lt;1..&gt;288

/product="16S ribosomal RNA"

77 a 61 c 88 g 62 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 85.5%; Score 18.8; DB 2; Length 288;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacagatttgata 22

Db 141 TATTGGATCAGGATTGATA 162

RESULT 8

LIU94978

LOCUS

Leptospira inadai 288 bp DNA BCT 01-JAN-1998

DEFINITION

Leptospira inadai 16S ribosomal RNA gene, partial sequence.

ACCESSION

U94978

VERSION

U94978.1 GI:2735450

KEYWORDS

SOURCE

Leptospira inadai.

ORGANISM

Bacteria; Spirochaetales; Leptosiraceae; Leptospira.

REFERENCE

1 (bases 1 to 288)

AUTHORS

Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M., Brenner, D.J. and Patel, B.K.C.

TITLE

Identification of Leptospira inadai by continuously monitoring

JOURNAL

fluorescence during rapid cycle PCR

REFERENCE

2 (bases 1 to 288)

AUTHORS

Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M., Brenner, D.J. and Patel, B.K.C.

TITLE

Direct Submission

JOURNAL

Submitted (24-MAR-1997) School of Science, Griffith University, Brisbane, QLD 4111, Australia

FEATURES

Location/Qualifiers

source

1. .288

/organism="Leptospira inadai"

/strain="218/95"

/db\_xref="taxon:29506"

&lt;1..&gt;288

/product="16S ribosomal RNA"

77 a 61 c 88 g 62 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 85.5%; Score 18.8; DB 2; Length 288;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacagatttgata 22

Db 141 TATTGGATCAGGATTGATA 162

RESULT 9

LIU94979

LOCUS

Leptospira inadai 288 bp DNA BCT 01-JAN-1998

DEFINITION

Leptospira inadai 16S ribosomal RNA gene, partial sequence.

ACCESSION

U94979

VERSION

U94979.1 GI:2735451

KEYWORDS

SOURCE

Leptospira inadai.

ORGANISM

Bacteria; Spirochaetales; Leptosiraceae; Leptospira.

REFERENCE

1 (bases 1 to 288)

AUTHORS

Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M., Brenner, D.J. and Patel, B.K.C.

TITLE

Identification of Leptospira inadai by continuously monitoring

JOURNAL

fluorescence during rapid cycle PCR

REFERENCE

2 (bases 1 to 288)

AUTHORS

Woo, T.H.S., Smythe, L.D., Symonds, M., Norris, M., Dohnt, M., Brenner, D.J. and Patel, B.K.C.

TITLE

Direct Submission

JOURNAL

Submitted (24-MAR-1997) School of Science, Griffith University, Brisbane, QLD 4111, Australia

FEATURES

Location/Qualifiers

source

1. .288

/organism="Leptospira inadai"

/strain="1078 VRI"

/db\_xref="taxon:29506"

&lt;1..&gt;288

/product="16S ribosomal RNA"

77 a 61 c 88 g 62 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 85.5%; Score 18.8; DB 2; Length 288;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacagatttgata 22

Db 141 TATTGGATCAGGATTGATA 162

RESULT 10

AB007012

LOCUS

Spirochaeta sp. 353 bp DNA BCT 13-OCT-1997

DEFINITION

Spirochaeta sp. 16S rRNA gene, partial sequence.

ACCESSION

AB007012

VERSION

AB007012.1 GI:2516253

KEYWORDS

SOURCE

Spirochaeta sp.

ORGANISM

Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.

REFERENCE

1 (bases 1 to 353)

AUTHORS

Shin, M.-S.

TITLE

Direct Submission

JOURNAL

Submitted (05-SEP-1997) to the DDBJ/EMBL/GenBank databases. Mi-Sun Shin, Laboratory of Marine Molecular Microbiology, Faculty of Agriculture, Kyoto University; Oiwakecho, Kitashirakawa, Sakyo-ku, Kyoto, Kyoto 306-01, Japan (E-mail: misunekais.kyoto-u.ac.jp, Tel: 075-753-6224, Fax: 075-753-6226)

REFERENCE

2 (sites)

AUTHORS

Shin, M., Yoshinaga, I., Uchida, A. and Ishida, Y.

TITLE

Phylogenetic analysis by 16S rRNA gene sequencing of Obligate

Oligotrophs isolated from the northern basin of Lake Biwa

(Mesotrophic Lake)

Unpublished (1997)

JOURNAL



Arabidopsis thaliana chromosome II section 214 of 255 of the complete sequence. Sequence from clones T517.	Arabidopsis thaliana
AC003000 AE002093	Arabidopsis thaliana
AC003000.2 GI:6598383	Arabidopsis thaliana
HTG.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana
thale cress.	Arabidopsis thaliana



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 109723)

AUTHORS Schudy,A., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K., Rump,A., Schillabel,M.B., Taudien,S., Wen,G., Schlegelberger,B., Siebert,R. and Rosenthal,A.

TITLE Chromosome 8 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 109723)

AUTHORS Schudy,A., Blechschmidt,K., Schillabel,M., Baumgart,C., Menzel,U., Weber,J., Schattevoy,R. and Rosenthal,A.

TITLE Direct Submission

JOURNAL Submitted (06-JUL-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany

REFERENCE 3 (bases 1 to 109723)

AUTHORS Genome Sequencing Center Jena.

TITLE Direct Submission

JOURNAL Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

COMMENT On Jun 1, 2000 this sequence version replaced gi:5514637.

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 61961: contig of 61961 bp in length

\* 61962 62061: gap of unknown length

\* 91746 91746: contig of 29685 bp in length

\* 91747 91846: gap of unknown length

\* 91847 109723: contig of 17877 bp in length.

FEATURES

Location/Qualifiers

1..109723

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="8 map 8q12-8q13"

/clone="CTA-397H3"

1..61961

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left"

91647..109723

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right"

31869 a 22757 c 23188 g 31709 t 200 others

BASE COUNT

ORIGIN

Query Match 85.5%; Score 18.8; DB 72; Length 109723;

Best Local Similarity 90.9%; Pred. No. 82;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttggtacacagatttgata 22

|||||

DB 95330 TTTTGGATCACAAAGATTGGTA 95351

RESULT 14

AF137379

LOCUS AF137379 200799 bp DNA circular PLN 14-SEP-1999

DEFINITION Nephroselmis olivacea chloroplast DNA, complete genome.

ACCESSION AF137379

VERSION AF137379.1 GI:5880684

KEYWORDS

SOURCE Nephroselmis olivacea.

ORGANISM Chloroplast Nephroselmis olivacea

Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Chlorodendraceae; Chlorodendraceae; Nephroselmis.

REFERENCE 1 (bases 1 to 200799)

AUTHORS Turmel,M., Otis,C. and Lemieux,C.

TITLE The complete chloroplast DNA sequence of the green alga Nephroselmis olivacea: insights into the architecture of ancestral chloroplast genomes

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (18), 10248-10253 (1999)

MEDLINE 99398694

REFERENCE 2 (bases 1 to 200799)

AUTHORS Turmel,M., Otis,C. and Lemieux,C.

TITLE Direct Submission

JOURNAL Submitted (24-MAR-1999) Biochimie, Universite Laval, Pavillon Charles-Eugene Marchand, Quebec G1K 7P4, Canada

FEATURES

Location/Qualifiers

1..200799

/organism="Nephroselmis olivacea"

/organelle="plastid:chloroplast"

/db\_xref="taxon:31312"

/cell\_line="NIES 484"

1..92126

/note="Region: large single-copy region (LSC)"

55..137

/gene="trnL(gag)"

/note="codons recognized: CUY"

/product="tRNA-Leu"

/anticodon="(pos:88..90,aa:Leu)

55..137

/gene="trnL(gag)"

332..634

/gene="rps14"

332..634

/gene="rps14"

/codon\_start=1

/product="ribosomal protein S14"

/protein\_id="AAD54778.1"

/db\_xref="GI:5880685"

/translation="MAKSMIERESKRAALVAKYATRRQALKAAIQKTSFDERLVLQ

HQLQDLPVNSVPCRLHNRCTITGRPKGYRDFGLSRHELRAHGCLLPGVTKASM"

765..849

/gene="trnL(uaa)"

/note="codon recognized: UUA"

/product="tRNA-Leu"

/anticodon="(pos:798..800,aa:Leu)

765..849

/gene="trnL(uaa)"

983..1054

/gene="trnQ(uug)"

/note="codons recognized: CAR"

/product="tRNA-Gln"

/anticodon="(pos:1015..1017,aa:Gln)

983..1054

/gene="trnQ(uug)"

1102..1182

/gene="trnL(uag)"

/note="codons recognized: CUR"

/product="tRNA-Leu"

/anticodon="(pos:1136..1138,aa:Leu)

1102..1182

complement(1364..1468)

/gene="psbM"

complement(1364..1468)

/gene="psbM"

/codon\_start=1

/product="M protein of photosystem II"

/protein\_id="AAD54779.1"

/db\_xref="GI:5880686"

/translation="MEVNIILGIATALFIIPTSPULLILYVKTSQQN"

1618..1691

/gene="trnM(cau)"

/note="codon recognized: AUG"

/product="tRNA-Met"

/anticodon="(pos:1652..1654,aa:Met)

1618..1691

/gene="trnM(cau)"

1733..3862

```

CDS
/gene="ftsI"
/1733..3862
/gene="ftsI"
/codon_start=1
/product="putative peptidoglycan synthetase"
/protein_id="AAD54780.1"
/db_xref="GI:5880687"
/translation="MIVNPMKYRROFRRLPKHGLIYVAFISISMIKIALDPYSKR
PKWMSGKPPROFENERMWMKLSLATVGLFATRLSGLOFNKYTELKSVAEQQIIGQ
VNDPQQRKILDHGDIVAIDLPAYDLVHPKMSLSLRIAEELSPILDSLOYLYN
RLDEGSGCICLHQIDTNSAQIRUGVGDGIELVHPQRYPKRGSFESILGVITDEG
YGOALESDDMMKSTYQDPCWMDGHGFLGIRFPKQILFHQESALQITIDCGLOE
KVSQLTNNAMRRGAKRIAI IEAHSGAIRCLATSPYDPCNGYGFPMFERCWPIT
DLFPGSTFKPNALALENGIFQPTDRILDITGKIRIGDSWIGVGGGTWDRSLDHL
TGTOILQRSSNVGMVMSLODPAIYHRNLIRLGLSHRNDNOTSEKMSHDSHNEGWL
NLADLSVDAISVVKDDQDEVDHEIEAATASFGQGLAMPKLLQILIAITANGMAVT
PHLSIKIVILDFHHQLHNSNEFSLOGVSVLSRQYHAKQRPRTYHLDYLGHPVP
SLELWGFVKSIPPTFRKRRFLSRQTCNVLLQGVLDQAQATSGSLFPGFANAG
KTGAQKASALGYSYDSVVTSEVGIYPAVKPFVTLVIIDEPDPFRFGFNAVDVT
QTLISEMIVQEQPPSPYTVLSLERN"
3998..5059
/gene="psbA"
/3998..5059
/gene="psbA"
/codon_start=1
/product="D1 reaction-center protein of photosystem II"
/protein_id="AAD54781.1"
/db_xref="GI:5880688"
/translation="MTAILRRETSVWARFCDMVTSTENRLYIGWFGVLMIPLLLTA
TSVPIIGFAAPVDVDIGIREPVSGSLICNNILISGAI IPSSAAIGIHFPYIWEASI
DEWLINGGCVELIVLHFLGACVIMGRENWELSRLOHMPWIAVAYAPVAAATAYELI
YPIQGQSGFDGMLIGISGTFNMFVFOAEHNLIMHPHMLGVAGVFGSGLFSAMHSSL
VTSLSIRRETTENASNAGYFGQEEETNIIVAAHYFGRLLFOYAFSNNRSRSLHFLA
AWPVVICWTFALGYSMAFNLNGFNFNQSVVDQGRVINTWADIINRANLGMVEVMHER
NANHPDLASVDAPAQVQ"
5489..6964
/gene="atpB"
/5489..6964
/gene="atpB"
/codon_start=1
/product="CF1 beta subunit of ATP synthase"
/protein_id="AAD54782.1"
/db_xref="GI:5880689"
/translation="MSQAGVKTKQAGKIVQIVGPMVDAVQPGSMPIYNALIVKGRN
NAGEVSVQCEVQOGLDGLVRAVMSATDGLMRGMEVTDTRALSVPVGPPTLGRIF
NVLEDPVNGPVNKEKTLTHREAPAVDLDTKLSIFETGIKVDLPLALYRRGGRTG
LFGAGVGKTVLIMELINNAKAGHGVFGVGGERTREGNDLYMEMCESKYRNKADY
KSSKVALVYQGMNEPPCARMKRVGLTALTMAEYFRDVNNQDVLFDINIFRFVQAGSEV
SALLGRMPSAVGQPTLATFEMGLOERITSTKDGSLTSIOAVYVPADDLTDPAATYF
AHLDATVLSRNLAAKGIYPADVPDLSTSTMLQPWILGDHDYGTQARVKOTLQRYKEL
ODITAILGDELSEEDRMIVARAKIERLUSQPPFVAEFTGAPKIVPLAESIQGFK
LILSGELSLPESAFYLVGNIEEAIAKAAASIQAAAAA"
6977..7378
/gene="atpE"
/6977..7378
/gene="atpE"
/codon_start=1
/product="CF1 epsilon subunit of ATP synthase"
/protein_id="AAD54783.1"
/db_xref="GI:5880690"
/translation="MSLQVRLLTPQVFLNVADEIILPTNTGOMGVLTNHTPLITAI
DIGPMLVRSSTWQSMALLGGLALVKDNOVITLVNEAELGSDINAEAEATTFLAAKEA
LANSKTRDQIENNAFRAKRVYQATLVK"
complement(7622..8104)
/gene="petB"
/complement(7622..8104)
/gene="petB"
/codon_start=1
/product="subunit IV of cytochrome b6/f complex"
/protein_id="AAD54784.1"
/db_xref="GI:5880691"
/translation="MSVTKKPDLTDPVLRAKLAKGMGHNYGEPAPWNDLLMYPVVI
LGTLSCTGLAVLDPAAIGEPANFATPLEILPEWYFFPVQLLRTPVKNLLGVLAA

AVPAGLLTVPFIESINKFQNPFRPVATTVFILGTVVATWALGIGATLPIDISLTGLGF"
gene
complement(8132..8779)
/gene="petB"
CDS
complement(8132..8779)
/gene="petB"
/codon_start=1
/product="apocytochrome b6 of cytochrome b6/f complex"
/protein_id="AAD54785.1"
/db_xref="GI:5880692"
/translation="MSKIYDWEERLEIOAIADDTSKYVPPHVNIFYCFGGTITLTCF
LIQVATFAMTYIRPTVEAFASVEYIMTNVFNGLIRSIHRWSAMVMMLILHVF
RYLVTGGFKPRELMTVGTILAVITVSEFVGYSLPMDQVGTWAKVITVGPDAIPV
IGAPLVELLRGVSVGQSTLTFRYSLSLHTEFLVPLLTAVFMLHFLMIRKOGISGPL"
complement(9045..9431)
/gene="rpl12"
/complement(9045..9431)
/gene="rpl12"
/codon_start=1
/product="ribosomal protein L12"
/protein_id="AAD54786.1"
/db_xref="GI:5880693"
/translation="MSTITNQIIEOLKSMTLLEAAELVSOIETFCVDASAPAAGAVM
VAAAPSPVVEEQTEFTVINDVPSAKRIAVIKVVRALTGLCLKEAKDMIESVPKAI
KEGIGKDEAQOIKQOLEEAGATVTK"
complement(9442..9846)
/gene="rps9"
/complement(9442..9846)
/gene="rps9"
/codon_start=1
/product="ribosomal protein S9"
/protein_id="AAD54787.1"
/db_xref="GI:5880694"
/translation="MDHTSSVGLRKTGRKCAVAQVRIIEGTQQLTINORPGVPLYOY
NPAYLLAAQCALDVTETLTRYDTIVTKGGTLTGQAEIKLGLARALCSLHIKRRAL
KPGYLTRNPLRKERYGLKAKKAPQFSKR"
complement(9998..11485)
/gene="rpoA"
/complement(9998..11485)
/gene="rpoA"
/codon_start=1
/product="alpha subunit of RNA polymerase"
/protein_id="AAD54788.1"

Query Match 85.58; Score 18.8; DB 6; Length 200799;
Best Local Similarity 90.94; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttttgatcacacaagattgata 22
||||| ||||| ||||| |||||
Db 134224 TGTGGCTCACATGATTGATA 134245

RESULT 15
AF137379/c AF137379 200799 bp DNA circular PLN 14-SEP-1999
LOCUS Nephroselmis olivacea chloroplast DNA, complete genome.
DEFINITION Nephroselmis olivacea chloroplast DNA, complete genome.
ACCESSION AF137379
VERSION AF137379.1 GI:5880684
KEYWORDS Nephroselmis olivacea.
SOURCE Nephroselmis olivacea.
ORGANISM Chloroplast Nephroselmis olivacea
Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;
Chlorodendreales; Chlorodendraceae; Nephroselmis.
REFERENCE 1 (bases 1 to 200799)
AUTHORS Turmel,M., Otis,C. and Lemieux,C.
TITLE The complete chloroplast DNA sequence of the green alga
Nephroselmis olivacea: insights into the architecture of ancestral
chloroplast genomes
Proc. Natl. Acad. Sci. U.S.A. 96 (18), 10248-10253 (1999)
MEDLINE 99398694
REFERENCE 2 (bases 1 to 200799)
AUTHORS Turmel,M., Otis,C. and Lemieux,C.

```

TITLE Direct Submission  
JOURNAL Submitted (24-MAR-1999) Biochimie, Universite Laval, Pavillon  
Charles-Eugene Marchand, Quebec G1K 7P4, Canada

FEATURES  
Location/Qualifiers  
1. 200799  
/organism="Nephroselmis olivacea"  
/organelle="plastid:chloroplast"  
/db\_xref="taxon:31312"  
/cell\_line="NIES 484"  
1. 92126  
/note="Region: large single-copy region (LSC)"

misc\_feature  
55..137  
/gene="trnL(gag)"  
/note="codons recognized: CUY"  
/product="trNA-Leu"  
/anticodon=(pos:88..90,aa:Leu)  
55..137  
/gene="trnL(gag)"  
332..634  
/gene="rps14"  
332..634  
/gene="rps14"  
/codon\_start=1  
/product="ribosomal protein S14"  
/protein\_id="AAD54778.1"  
/db\_xref="GI:5880685"  
/translation="MARKSMIERESKRAALVAKYATRRQALKAAIQKTSFDERLVQ  
HQQLPVSVPCLHNRCTTGRPKGYRDFGLSRHRLAMAGHCLLPVTRASW"

trNA  
765..849  
/gene="trnL(uaa)"  
/note="codon recognized: UUA"  
/product="trNA-Leu"  
/anticodon=(pos:798..800,aa:Leu)  
765..849  
/gene="trnL(uaa)"  
983..1054  
/gene="trnQ(uug)"  
/note="codons recognized: CAR"  
/product="trNA-Gln"  
/anticodon=(pos:1015..1017,aa:Gln)  
983..1054  
/gene="trnQ(uug)"  
1102..1182  
/gene="trnL(uag)"  
/note="codons recognized: CUR"  
/product="trNA-Leu"  
/anticodon=(pos:1136..1138,aa:Leu)  
1102..1182  
/gene="trnL(uag)"  
complement(1364..1468)  
/gene="psbM"  
complement(1364..1468)  
/gene="psbM"  
/codon\_start=1  
/product="P protein of photosystem II"  
/protein\_id="AAD54779.1"  
/db\_xref="GI:5880686"  
/translation="MEYNILGLIATLFIITPSFLILYVKTASQON"  
1618..1691  
/gene="trnM(cau)"  
/note="codon recognized: AUG"  
/product="trNA-Met"  
/anticodon=(pos:1652..1654,aa:Met)  
1618..1691  
/gene="trnM(cau)"  
1733..3862  
/gene="ftsI"  
1733..3862  
/gene="ftsI"  
/codon\_start=1  
/product="putative peptidoglycan synthetase"  
/protein\_id="AAD54780.1"  
/db\_xref="GI:5880687"

/translation="MIYNPMKYRRQRRLPKHLQGIYYAFSLSTMIKIKIALDPSKR  
PMKMHSGTFFQYENRVMIKSLATVGLLFATRLSGLOFNKYTELKSVAEQOIGQ  
VNDQQRKIIILDHGDIVADLPAYDLYVHPRMCSISLERIAELLSPLDLSQOYL  
RLDEGDSICLMOHIDTNTSAQIRRLGVGDIELVHPQVVPKRGSESLGYVDTG  
YGOAGLESSLDDNMKSTYQDVPCMDGNGFLGIRPKQILFHQESALQLTIDCLOE  
KVSOLITNAMNRFCAKRIAIIMEAHSGARCLATSPSYDNCYGVWPFMERPCWPI  
DLFEPGSTFKPVLNIALENGIIFOPTDRILDITGIRIGDSWGNVGGFIDWRLDHL  
TGTOILQORSNVGMVRYMQSLDPAIYHRNLIRLGLGRNDNOTSFKMSSHDHNSGW  
NLKDLTSDYAIISVVKDQDEFVDHEIEATASFQGLAMTPLKLLQLLATANGCMVT  
PHLTSKIVTLDFHHLHLMNEFSLQMGOSVLSRSOYHAKQPRPYTHLYLGHVPVP  
SLELGFWDVKSIPHTERRRLFSRQTCNVLLGNLEQVLDAAQATSGRFLPGVAMG  
KTGTAKASALGYSYSDSVTSFVGIYPAVKPKFVTLVIDEPPDFRGENTAVDT  
QTLLISEMIVQEQDPPSTPTVSLFERNM"  
3998..5059  
/gene="psbA"  
3998..5059  
/gene="psbA"  
/codon\_start=1  
/product="D1 reaction-center protein of photosystem II"  
/protein\_id="AAD54781.1"  
/db\_xref="GI:5880688"  
/translation="MTAILERRESTSVARFCDMWTSTENRLYIGWFGVLMIPLLLTA  
TSVFIIGFIAAPPVDIDGIREPVSGLLFGNNIISGAIIPSSAAIGIFHYPIWEAASI  
DEWLYNGCYELIVLHLLGVACVNGREWELSPRLGMRPWITAYASAPAAATVFLI  
YPIGQSFSDGMPGLGISGTENFMIVQAEHNLIMHPFHMUGVAVFGSGLFSAMHGS  
VTSLLIETTESANAGYKFGQEEETYNIVAAHGYPFGLIFQYAFSNNSSRLSHFLA  
AWPVYCITWFTALGVSTMAFNLNGFNQSVVDSQVINTWADIINRANLGMVWHER  
NAHNFLLDASVDAPAVQG"  
5489..6964  
/gene="atpB"  
5489..6964  
/gene="atpB"  
/codon\_start=1  
/product="CF1 beta subunit of ATP synthase"  
/protein\_id="AAD54782.1"  
/db\_xref="GI:5880689"  
/translation="MSQAGVYTKQAGKIVQIVGPMVDAFQPGMPNINLIYKGRN  
GAGQEVSVCEVQQLGDLVRAVMSATGDLMRGMEVDTGRALSPVGPVGTIGRIF  
NVLGEVDNMGVGNKTLPIHREAPAFVLDLTKLSTFETGKYVDVLGPPVRRGKIG  
KFGGAGVGYVLMELINNTAKAHGVSFGVGERTNGDLYNMENCESKVIINKADV  
KSSKVALYQOMNEPPGARMRVGLTALTMAEYFRDVANNODVLLFDNIERFVQAGSEV  
SALIGRMPASVGYOPTLATENGLOERTSTKDGSIISQAVYVPADLTDPAATTF  
AHLDTTVLSRNLAAKGIYPADVPLDSTSLMQPWLGDHGHYGAQRVKQLRYKEL  
QDIITAILGDELSEEDMIVARAKKIERLSPFVAEFTGAPGKYVPLAESIQGKEL  
LILSGELDSLPEAFYLVGNTEEAIAKASIAQAAAAA"  
6977..7378  
/gene="atpB"  
6977..7378  
/gene="atpB"  
/codon\_start=1  
/product="CF1 epsilon subunit of ATP synthase"  
/protein\_id="AAD54783.1"  
/db\_xref="GI:5880690"  
/translation="MSLQVRLLTPVOVLNVSADEILPTNTGOMGVLTNHTPLITAI  
DIGPMLYRSESTWQSMALLGGLALVKNQVILLVNEAELGSDINAEAEETFLAAKE  
LANSTRKQDIENNLAFKARVRYQATLVK"  
complement(7622..8104)  
/gene="petD"  
complement(7622..8104)  
/gene="petD"  
/codon\_start=1  
/product="subunit IV of cytochrome b6/f complex"  
/protein\_id="AAD54784.1"  
/db\_xref="GI:5880691"  
/translation="MSVTKKPDLTDPVLRAKLAKMGHNYGGEPAWPNLDLIMFPVVI  
LGTLSCTIGLAVLDPAAGEPANPEATPEILPEWYFPVQLLRTVPNKLGLVLLMA  
AVPAGLLTVFIESINKFQNPFRPVATVFLIGTVVAILGLGATLFDISLTLGLF"  
complement(8132..8779)  
/gene="petB"  
complement(8132..8779)  
/gene="petB"  
/codon\_start=1

```
/product="apocytochrome b6 of cytochrome b6/f complex"
/protein_id="AAD54785.1"
/db_xref="GI:5880692"
/translation="MSKIYDFEERLEIQAIADDTTSKYVPHVNIIFYCFGGITLTCF
LIQVATGFAMTYFRVTVEAFASVEYIMTNVFWGLIRSIHRWSASMMVMLLHV
RVLTGGFKPRELTVGTGVIILAVITVSEFVGYSLPQDQGVYAWKIVTGPDAIPV
IGAPLVLLRGSVSGQSTLIRFYSLHTFVLPLLTAVFMLHFLMIRKQGISGPL"
complement(9045..9431)
/gene="rpl12"
complement(9045..9431)
/gene="rpl12"
/codon_start=1
/product="ribosomal protein L12"
/protein_id="AAD54786.1"
/db_xref="GI:5880693"
/translation="MSTITNQIIIEOLKSMTLLEAEELVSOIETFGVDASAPAAGAVM
VAAAPSAFVVEQTEFTVINDVPSAKRIAVIKVVRALTGLGLKEAKDMIESVPKAI
KEGIGKDEAQQINQQLLEEAGATVTR"
complement(9442..9846)
/gene="rps9"
complement(9442..9846)
/gene="rps9"
/codon_start=1
/product="ribosomal protein S9"
/protein_id="AAD54787.1"
/db_xref="GI:5880694"
/translation="MDHTSSVGLRKTGRKKCAVAQVRIIEGTGQLTINORPGVPYIQY
NPAYLLAAGCALDVTELTNRYDTIVKTKGGLTGQAEAIKGLARALCSLHIKHKRAL
KPGYLTRNPLRKRKKYGLKKARKAPQFSKR"
complement(9998..11485)
/gene="rpoA"
complement(9998..11485)
/gene="rpoA"
/codon_start=1
/product="alpha subunit of RNA polymerase"
/protein_id="AAD54788.1"
```

```
Query Match      85.5%  Score 18.8; DB 6; Length 200799;
Best Local Similarity 90.9%  Pred. No. 77;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 tgttgatcacagaatttgata 22
      ||||| ||||| ||||| |||||
Db 158702 TGTGGCTCACATGATTGATA 158691
```

Search completed: March 27, 2001, 08:17:03  
Job time: 5900 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 06:38:43 ; Search time 2286.42 Seconds  
(without alignments)  
3306.005 Million cell updates/sec

Title: US-09-380-826A-1  
Perfect score: 1477  
Sequence: 1 gatcatggctcgaactaac.....ccgtaaatcgtctctgcag 1477

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_ph.\*
- 6: gb\_pl1.\*
- 7: gb\_pl2.\*
- 8: gb\_pr1.\*
- 9: gb\_pr2.\*
- 10: gb\_pr3.\*
- 11: gb\_ro.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: em\_fun.\*
- 15: em\_hum1.\*
- 16: em\_hum2.\*
- 17: em\_in.\*
- 18: em\_om.\*
- 19: em\_or.\*
- 20: em\_ov.\*
- 21: em\_pat.\*
- 22: em\_ph.\*
- 23: em\_pl.\*
- 24: em\_ro.\*
- 25: em\_sts.\*
- 26: em\_sy.\*
- 27: em\_un.\*
- 28: em\_v1.\*
- 29: gb\_ba3.\*
- 30: gb\_in1.\*
- 31: gb\_in2.\*
- 32: gb\_in3.\*
- 33: gb\_pl3.\*
- 34: gb\_pl4.\*
- 35: em\_ba1.\*
- 36: em\_ba2.\*
- 37: em\_htg1.\*
- 38: em\_htg2.\*
- 39: em\_htg3.\*
- 40: em\_htg4.\*
- 41: em\_htg5.\*
- 42: em\_htg6.\*
- 43: em\_htg7.\*

- 44: em\_htg8.\*
- 45: em\_htg9.\*
- 46: em\_htg10.\*
- 47: em\_hum3.\*
- 48: em\_hum4.\*
- 49: em\_hum5.\*
- 50: em\_hum6.\*
- 51: gb\_pr5.\*
- 52: gb\_pr6.\*
- 53: gb\_pr7.\*
- 54: gb\_htg1.\*
- 55: gb\_htg2.\*
- 56: gb\_htg3.\*
- 57: gb\_htg4.\*
- 58: gb\_htg5.\*
- 59: gb\_htg6.\*
- 60: gb\_htg7.\*
- 61: gb\_htg8.\*
- 62: gb\_htg9.\*
- 63: gb\_htg10.\*
- 64: gb\_htg11.\*
- 65: gb\_htg12.\*
- 66: gb\_htg13.\*
- 67: gb\_htg14.\*
- 68: gb\_htg15.\*
- 69: gb\_htg16.\*
- 70: gb\_htg17.\*
- 71: gb\_htg18.\*
- 72: gb\_htg19.\*
- 73: gb\_htg20.\*
- 74: gb\_htg21.\*
- 75: gb\_htg22.\*
- 76: gb\_htg23.\*
- 77: gb\_sts1.\*
- 78: gb\_sts2.\*
- 79: gb\_v1.\*
- 80: gb\_v12.\*
- 81: gb\_pat1.\*
- 82: gb\_pat2.\*
- 83: em\_htg0.\*
- 84: gb\_htg24.\*
- 85: gb\_pr8.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description
1	1431.4	96.9	1481	2	LFU60594	U60594	Leptosira
2	1388.6	94.0	1450	2	LFA19243	Y19243	Leptosira
3	1295.6	87.7	1508	2	LIRRN16S	X17547	Leptosira
4	1270	86.0	1374	2	L116SRDNY	Z21634	L.inadai ge
5	1263.8	85.6	1483	29	WA4516S	U12676	Leptosira
6	1263.8	85.6	1489	29	WA5216S	U12677	Leptosira
7	1262.2	85.5	1513	2	LBUI12670	U12670	Leptosira
8	1261	85.4	1489	2	LBUI12669	U12669	Leptosira
9	1260.8	85.4	1516	2	LNUI12671	U12671	Leptosira
10	1254.6	84.9	1425	2	LB16SRDNP	Z21630	L.borgpeter
11	1251.8	84.8	1486	29	WB4616S	U12673	Leptosira
12	1246.6	84.4	1419	2	LK16SRDNP	U12672	L.kirschner
13	1239.8	83.9	1494	2	LSU12672	U12672	Leptosira
14	1238.2	83.8	1418	2	LM16SRDNY	Z21648	L.meyeri ge
15	1232.8	83.5	1396	2	L116SRRN	Z21817	L.interroga
16	1230	83.3	1415	2	LS16SRDNY	Z21649	L.santarosa
17	1225.4	83.0	1410	2	LN16SRDNP	Z21637	L.weilli ge
18	1216	82.3	1424	2	LN16SRDNP	M88721	Leptosira
19	1063.4	72.0	1523	2	LEPRR16SB	Z98590	Leptosira
20	1031.8	69.9	1433	2	LFTM61993	Z21638	L.wolbachii
21	1027.6	69.6	1417	2	LN16SRDNY		







QY	1281	aagtcggaatcgttagtaatcgcgatacagcatccgcggtgaatacgttcccgacctt	1340
Db	1261	AAGTCGGAATCGCTAGTAATCGCGGATCAGCATGCCGGGTGATACGTTCCCGGACCTT	1320
QY	1341	gtacacaccgcgtcacaccacttgatggtgggagcaccgaaagtgcctttgttaacg	1400
Db	1321	GTACACACCGCCGTCACACCACCTGAGTGGGAGCACCAGTGGTCTTTGTTAAACG	1380
QY	1401	taaggagacagactactaaagttaaaactcgtataaagggtgaagtcgtaacaaggtacc	1459
Db	1381	TAAGGAGACAGACTACTAAAGTGAACCTCGTAAAGGGGTGAAGTCGTAAAGGTAGC	1439
RESULT	3		
LIRN16S		LIRN16S 1508 bp DNA BCT 04-SEP-1991	
LOCUS		Leptospiira interrogans 16S ribosomal RNA gene.	
DEFINITION		X17547	
ACCESSION		X17547.1 GI:44008	
VERSION		16S ribosomal RNA; ribosomal RNA.	
KEYWORDS		Leptospiira interrogans.	
SOURCE		Leptospiira interrogans.	
ORGANISM		Bacteria; Spirochaetales; Leptospiiraceae; Leptospiira.	
REFERENCE		1 (bases 1 to 1508)	
AUTHORS		Fukunaga, M.	
TITLE		Direct Submission	
JOURNAL		Submitted (22-NOV-1989) Fukunaga M., Faculty of Pharmacy and Pharmaceutical Science, Fukuyama University, 985 Higashimuracho, Fukuyamashi, Hiroshimaken 729-0, Japan	
REFERENCE		2 (bases 1 to 1508)	
AUTHORS		Fukunaga, M., Horie, I., Okuzako, N. and Mifuchi, I.	
TITLE		Nucleotide sequence of a 16S rRNA gene for Leptospiira interrogans serovar canicola strain Moulton	
JOURNAL		Nucleic Acids Res. 18 (2), 366 (1990)	
MEDLINE		90221824	
COMMENT		Data kindly reviewed (23-FEB-1990) by Tukunaga M.	
FEATURES		Location/Qualifiers	
source		1..1508	
tRNA		/organism="Leptospiira interrogans"	
		/strain="serovar canicola strain Moulton"	
		/db_xref="taxon:173"	
		1..1508	
		/note="16S ribosomal RNA"	
BASE COUNT		393 a 342 c 456 g 317 t	
ORIGIN			
Query Match		87.7%; Score 1295.6; DB 2; Length 1508;	
Best Local Similarity		94.7%; Pred. No. 5.4e-14;	
Matches 1384; Conservative		0; Mismatches 14; Indels 4; Gaps	
QY	1	gatcatggctcagaactaacctgcgcgcgctttaaactcgaatcgaatcgagcggttag	60
Db	15	GATCTGGCTCAGAACTAACGCTGGCGCGGCTCTTAAACATGCACTCAAGCGGAGTAG	74
QY	61	caataactagcgcgaacgggtgagtaacacgt-ggtaactcttcccgagctcgagtag	119
Db	75	CAATCTACCGCGNACGGGTGAGTAAACACGTGGGTAACTCTCTCTGAGTCFGGGATA	134
QY	120	acttccgaagaagactaatcacgtagtcctctgttggatccacaagatttgatagta	179
Db	135	ACTTTCGGAAGGGAAGCTAATCTGATGTCCTCCGAGAGATCACAAGATTTTTCGGGTA	194
QY	180	aagatttattgcttggagatgagccgcgcgcgattagctagttgtaggttaatggctc	239
Db	195	AAGATTTATTCTCGGAGATGAGCCCGCGCTCCGATTAGTCTTGGTGGGTAAAGGCTC	254
QY	240	accaaggcagatcgtatgacccgcctgagaggtgtccggccacaatggaaactgagaca	299
Db	255	ACCAAGGCGAGTCGTGATGCCGCTTGACAGGGGTGTTCCGGCACAAATGGAACTGAGACA	314
QY	300	cgggtccatactctacgaggagcgaggttaagaatcttgctcaatgggggaaaccctga	359

Db	315	CGTCCATACTCTACGGGAGCGCAGCAGTTTAAAGAAATCTTGCTCAATGGGGGAA-CCTGA	373
Qy	360	agcagcgacgcgcgtgaacgaagaagctctcggattgtaaagttcattagca-ggaa	418
Db	374	AGCAGGAGCGCGGTGAACGATGAAGGCTCTTCGGATGTAAAGTTTCAGTAAGCAGGGAA	433
Qy	419	aaataagcagcaatgtgatgtgtacctgctaagaacaccgcgctaactacgtgcacga	478
Db	434	AAATAGCAGCAATGTGATGGTACCTTGCCTAAGCACCGGGTAACTACGTGCCAGCA	493
Qy	479	gccgcggttaacgtatgtatgcaagcgttgttcggaaatcattggcgctaaaggggtgcga	538
Db	494	GCCCGGTAAATACGATATGTGTGCAAGCGTTTTCGGAATCAITTTGGCGCTAAGAGGTGCGTA	553
Qy	539	ggcggaatttgtaagtcagggtgtgaaaacctgcgggctcaaccgcgtggcctgcacttgaac	598
Db	554	GGCGACATGTAACTGAGGTGTGAAGCTGCGGGCTCAACTGCGACGCTGCACTTGAAC	613
Qy	599	tacaagtcgagtttggagagcgcaagtggaaattccagagttagcggtagaaatgcgtag	658
Db	614	TATGTGCTCGGNGTTGGGAGAGCGCAAGTGGAAATTCACGGGTACGGGTGAATGCGGTAG	673
Qy	659	atatctggaggaaacaccagtgcgaagcgacttgtcgtctcaaaactgaagtcgagca	718
Db	674	ATATCTGAGAGAACACCAGTGGCGAAGCGCACTTCTGGCTAAAACTGACGCTGAGGCA	733
Qy	719	cgaagcgtgggttagtaaacgggattagatacccccgttaatccaccccctaagctgttc	778
Db	734	CGAAGCGTGGGTAGTGAACGGGATTAGATACCCCGGTAATCCACGCCCTTAACGTTGTC	793
Qy	779	taccagttgggggttttaaccctcagtaacgaacctcaacggaattaagttagaccgct	838
Db	794	TACAGTTGTGGGGGTGTTTACCCTCAGTAACGAACCTAACCGATTAAGTAGACGCCCT	853
Qy	839	gggactatgctgcgaagagtgaaactcaaggaattgaagggggtccgcacaagcggtg	898
Db	854	GGGACTATGCTGCGAAGAGTGAACCTCAAGAGGAATTACGGGGGTCCGCACAAGCGGTG	913
Qy	899	gagcatggtttaatttcgatgatatacccaaaaacctcacctggcctgacatgatctg	958
Db	914	GAGCATGTGGTTTAATTCGATGATACGCGAAAAACCTCACCTAGCTTGACNTGGAGTGG	973
Qy	959	aatactgtagagatatatgacctcggcagattcacagggtgctgcatggtctgcga	1018
Db	974	AATCATGTAGAGATACATGAGCCCTTGGGCCGCTTCACAGGTGCTGATGTTGCTCGTCA	1033
Qy	1019	gtcgtgctgtgagatgttgggttaagtcgccgcaagcagcaacccctatcgtatgtg	1078
Db	1034	GCTCGTGTGCGTGAATGTGGGTTTAAGTCCCGCAACGAGCGCAACCTCACCTTATGTTG	1093
Qy	1079	cta-ccctaagtggcactggtacgaactgcggttgacaacacggaggaaagcgggga	1137
Db	1094	CCATCAITTCAGTTGGGACTCGTAAGGAACCTGCGGTGACAAACCGGAGAGGCGGGGA	1153
Qy	1138	tgagtcgaactcctcatggcctttatgtccaggccacacacgtgtctaaatggccgata	1197
Db	1154	TGACGTCAAAATCCTCATGGCCCTTTATGTCTAGGGCAACACACGTGCTACAAATGGCGGTA	1213
Qy	1198	cagaggtgcgaactcgcaagggagctaaatctctaaagtcggtcccgatcgagtt	1257
Db	1214	CAAAGGTTACCAACTCGCGAGGGGAGCTAATCTCAAAAATCCGGTCCCGATTCGGATT	1273
Qy	1258	ggggctcgcaactcgaccccaatgaagtcggaatcgctagtaatcgcggaatcagcatgcgcg	1317
Db	1274	GGAGTCTGCAACTCGACCTCCATGAAGTCGGATCGCTAGTAATCGCGGATCAGCATGCCG	1333
Qy	1318	cggtgaaatagtttcccggaacttgtacacacgcgcgcgtcaacacactgagtgaggagca	1377
Db	1334	CGGTGATATGTTCCCGGAGCTTGTACACACGCCCGGTACACACCACTGAGTGGGAGCA	1393
Qy	1378	cccgagtggtcttttgttaaccgtlaaggagacagactactaaggtagaaactcgtaaaagg	1437
Db	1394	CCCGAAGTGGTCTTTTGCAACCCCAAGGAGCAGACTACTTAAGTGAACCTCGTGAAGGG	1453

```

Qy 1438 ggtgaagtcgttaacaaagtacc 1459
Db 1454 GGTGAAGTCGTAACAAGGTAGC 1475

RESULT 4
L116SRDNY
LOCUS L116SRDNY 1374 bp DNA BCT 02-DEC-1993
DEFINITION L.inadai gene for 16S ribosomal RNA (partial).
ACCESSION Z21634
VERSION Z21634.1 GI:433581
KEYWORDS 16S ribosomal RNA.
SOURCE Leptospira inadai.
ORGANISM Bacteria; Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE 1 (bases 1 to 1374)
AUTHORS Hooke, J.V.
TITLE Phylogeny of Leptospiaceae and related Spirochaetes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1374)
AUTHORS Hooke, J.V.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-1993) Hooke J.V., Public Health Laboratory
Service, Leptospira Reference Laboratory, Stonebow Road, Hereford,
Herefordshire, United Kingdom, HRI 2ER
FEATURES
source
1..1374
/organism="Leptospira inadai"
/strain="LYME"
/db_xref="taxon:29506"
<1..>1374
/rRNA
BASE COUNT 358 a 311 c 413 g 289 t 3 others
ORIGIN

Query Match 86.08; Score 1270; DB 2; Length 1374;
Best Local Similarity 96.58; Pred. No. 1.5e-13;
Matches 1327; Conservative 2; Mismatches 43; Indels 3; Gaps 3;

Qy 16 ctaacgtcgccggtcttaacatcgaagtcgagcggtgagcaatacctacgcgcg 75
Db 1 STARCGTTGGCGGCGCTCTTAACATCATCAAGTCGAGCGGGTAGCAATACCTACGCGCG 60

Qy 76 aacgggtgagtaaacagt-ggtaattcttccctccgagtcggtggataaactttccgaagaa 134
Db 61 AACGGGTGAGTAACACGTTGGTAATCTTCTCCGAGTCTGGGATAACTTTCCGAAGGAA 120

Qy 135 agctaataccggatagtcctgttgatcacagaatttatgaggttaagattattgcttg 194
Db 121 AGCTAATAACCGGATAGTCTACTGGATCACAGGATCTGATAGGTAAGATTATTGCTTG 180

Qy 195 gagatgagccgcccattagcttagttgtgaggttaagttaagctcaccaagcagcagtc 254
Db 181 GAGATGAGCCCGCCGATAGCTAGTTGTGAGTAAAGCTCACCAGCGACGATC 240

Qy 255 ggtagccgcttgagaggggtctccggccacaatggaactgagacacggtccatacctcta 314
Db 241 GGTAGCCGCTGAGAGGGTGTCCGGCCACAATGGAACGTGAGACACGCTCCATACTCCTA 300

Qy 315 cgggagcagcaggttaagaattctgtctcaatgggggaacacctgaagcagcagccgcg 374
Db 301 CGGGAGCAGCAGGTAAAGAACTTGTCTCAATGGGGGAACCTTGAAGCAGCGACGCCGCG 360

Qy 375 tgaacgaagaaggtcttcggtattgaaagttcattagca-ggaataataagcagcaatg 433
Db 361 TGAACGAAGAAGGCTCTCGGATTGTAAGCTCAATTAACGACGGAAAAATAGACCAATG 420

Qy 434 tgatgatggtacgttcctaaagcagcggcctaactacgtgcagcagccgcggtatacgt 493
Db 421 TGATGATGTTACCTGCCTAAAGCACCAGGCTAACACGTTGCGGAGCAGCGCGGTAATACGT 480

```

## RESULT 5

```

WA4516S
LOCUS Leptospira weilii 1483 bp DNA BCT 18-FEB-1995
DEFINITION Leptospira weilii Cellidoni (ATCC 43285) 16S rRNA gene, partial
sequence.
ACCESSION U12676
VERSION U12676.1 GI:558937
KEYWORDS Leptospira weilii.
SOURCE Leptospira weilii.
ORGANISM Bacteria; Spirochaetales; Leptospiaceae; Leptospira.
REFERENCE 1 (bases 1 to 1480)

```

AUTHORS Ralph,D. and McClelland,M.  
TITLE Phylogenetic evidence for horizontal transfer of an intervening  
sequence between species in a spirochete genus  
J. Bacteriol. 176 (19), 5982-5987 (1994)  
MEDLINE 95014031  
REFERENCE 2 (bases 1 to 1483)  
AUTHORS McClelland,M.  
TITLE Direct Submission  
JOURNAL Submitted (25-JUL-1994) Michael McClelland, California Institute of  
Biological Research, 11099 North Torrey Pines Road, La Jolla, CA  
92037, USA

FEATURES Location/Qualifiers  
source 1. .1483  
/organism="Leptospira weilii"  
/strain="serogroup Celledoni, serovar celledoni, strain  
Celledoni (ATCC 43285)"  
/isolate="WA45"  
/db\_xref="taxon:28184"  
<1. .1483  
/product="16S rRNA"  
BASE COUNT 382 a 331 c 452 g 315 t 3 others  
ORIGIN

Query Match 85.68; Score 1263.8; DB 29; Length 1483;  
Best Local Similarity 94.18; Pred. No. 1.7e-13;  
Matches 1376; Conservative 0; Mismatches 80; Indels 6; Gaps 6;

QY 1 gacatggtcgaactaacgctggcgccgctcttaaacatgcaagtcagcgggtag 60  
DB 10 GATCTGGCTCAGAACTAACCGTGGCGCGGCTCTTAACATGCAAGTCAAGCGCAGTAG 69  
QY 61 caatacctagcgccgaacgggtagtaacacagt-ggtaattctctcgaactcggata 119  
DB 70 CAATACTACGCGCAACGGGTAGTAACACGTGGGTAACTCTCTCCGAGATCTGGGATA 129  
QY 120 actttccgaagaagactaataccgagtagctctgttggatcaagaatttgatagta 179  
DB 130 ACTTTCGGAAGGAGGAGCTAATCTGATGGTGGTCCGAGAGGTCTATGATGTTTCGGGTA 189  
QY 180 aagatttattgttgagatgagcccgccgagtagctagctagttgtaggtaattggctc 239  
DB 190 AAGATTATTTGCTCGGAGCTGAGCCCGCCCGCTATGATGTTGGTGGGTAATGGCTC 249  
QY 240 accaagcgacatcgttagcgccctgaggggtgctccgcccacaaatggaactgagaca 299  
DB 250 ACCAAGCGAGATCGGTAGCCGGCTGAGAGGGTGTTCGGCCACAATGGAACCTGAGACA 309  
QY 300 cggctccatactctacggaggcagcagtagtaagaatctctgctcaatgggggaaacccctga 359  
DB 310 CGGTCCATACT-CTACGGGAGGAGCAGTAAAGAACTTTGCTCAATGGGGGAAACCCCTGA 368  
QY 360 agcagcgaccccggtgaacgaagaaggctctcgattgttaaagttcattagga-ggaa 418  
DB 369 AGCAGCAGCCCGGTGAACGATGAAGTCTTCGGATTGTAAGTTCAATAAGCAGGGAA 428  
QY 419 aataaagcagaatgtgatgtgacctgacctaaagcaccggctaacctacgtgacagca 478  
DB 429 AATAAGCAGCGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 488  
QY 479 gcccggttaatacgtatggtgcaagcgtttctcggaatcattggcgtaaaagggtgcgta 538  
DB 489 GCCCGGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 548  
QY 539 gccgatttgatcaggtgtgaaactcgccgctcaacccctggcctgacctgaaac 598  
DB 549 GCGGACATGTAACTCAGGTGTGAACACTCGGGCTCAACTCGCAGCTGCACCTTGAAC 608  
QY 599 tacaagtgaggtttgggagaggcaagtggaattccagggttagcgggtgaaatgcgtag 658  
DB 609 TATGTGCTGGAGTTGGGAGAGGCAAGTGAATTCAGGTGAGCGGTGAATGCGGTAG 668  
QY 659 atatctgagggaacaccagtggtggcgaaggcagctgtgctggtcctcaaacactgacgctgaggca 718

DB 669 ATATCTGGAGAACACACAGTGGCGAAGCGACATTGCTGGCTAAAACTGACGCTGAGGCA 728  
QY 719 cgaagcgtgggttagtaaaacggattagataccggtaataatccacgcctaaacgtgtgc 778  
DB 729 CGAAAGCGTGGGTAGTGAACGGATTAGATACCCCGGTAACTACACGCCCTTAACAGCTGTC 788  
QY 779 taccagttgtgggggttttaaccctcagtaacgaacctaacggattaaatagacgcct 838  
DB 789 TACCAGTTGTTGGGGTTTAAACCTCAGTAACGAA-CTAACGGATTAAAGTAGACCGCT 847  
QY 839 ggggactatgctcgaagagtgaactcaaaagggaatgacgggggttcgcacaacggtg 898  
DB 848 GGGGACTATGCTCGCAGAGTGAACACTCAAAAGGAATTGACGGGGTNNNCAACAGCGGTG 907  
QY 899 gacgatgggttaattcagatgatacccaaaacccctacccctggcttgacatggatctg 958  
DB 908 GAGCATGTGGTTTAAATTCGATGATACGCGAANAACCTCACCTAGCTTGACATGGAGTGG 967  
QY 959 aatcatgtagagatatagaccttcggcgagattcacaggtgctgcatggtgtctgctca 1018  
DB 968 AATTATGTAGAGATACATAGAG-CTTCGGGCGCTTTCACAGGTGCTGATGTTGTGTTCA 1026  
QY 1019 gctcgtgctgtagatgttggttaagtcccgcaacgagcgcaacccctatcgatgttg 1078  
DB 1027 GCTCGTCTGATGAGATGTTGGGTAAAGTCCCGCAACGAGCGCAACCTCACCCTTATGTTG 1086  
QY 1079 cta-cccttaagtgtggcactggtacgaactcgcgggtgacaaacccggaggagcgggga 1137  
DB 1087 CCATCATTTAGTTGGGCACCTCGTAAGGAACCTGCCGTGACAAACCGAGAGCGGGGA 1146  
QY 1138 tgacgtcaaatctcctatgaccttattgctcaggggcacaacagctgctacaaatggcgata 1197  
DB 1147 TGACGTCAAAATCTCATATGCTTATGCTAGGGCAACACAGTGTACAAATGGCCGTA 1206  
QY 1198 caagaggtgcgaactcgcgaagggagctaatctctaaagtcggtccagttcggtatt 1257  
DB 1207 CAAAGGGTAGCCAACTCGCGAGGGGAGCTAATCTCAAAAACCGCTCCAGTTCGGATT 1266  
QY 1258 ggggtctgaactcgaccctcgaatcggaatcggttagtaacgtcggtccagtcagatcgccg 1317  
DB 1267 GGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGTAGTAATCGCGGATCAGATGCCG 1326  
QY 1318 cgtggaatacgttcccgacctgttacacaccccgctcacacccctagtggtgggagca 1377  
DB 1327 CGGTGAATAGTTCCTCCGAGCTTGTACACACCGCGTACACCACTGAGTGGGAGCA 1386  
QY 1378 ccggaagtgtcttctgttaaccgttaaggagacagactactaaagtgaaactcgtaagg 1437  
DB 1387 CCCGAAGTGGTCTTTGCGCAACCGTAAGGAAGCAGACTTACTAAGGTGAAACTCGTGAAGG 1446  
QY 1438 ggtgaagtcgtaacaagggtacc 1459  
DB 1447 GGTGAAGTCGTAAACAGGTACC 1468

RESULT 6  
LOCUS WA5216S 1489 bp DNA BCT 18-FEB-1995  
DEFINITION Leptospira weilii Worsfold 16S rRNA gene, partial sequence.  
ACCESSION U12677  
VERSION U12677.1 GI:558938  
KEYWORDS  
SOURCE Leptospira weilii.  
ORGANISM Leptospira weilii  
REFERENCE 1 (bases 1 to 1480)  
AUTHORS Ralph,D. and McClelland,M.  
TITLE Phylogenetic evidence for horizontal transfer of an intervening  
sequence between species in a spirochete genus  
J. Bacteriol. (1994) In press  
REFERENCE 2 (bases 1 to 1489)  
AUTHORS McClelland,M.

TITLE Direct Submission  
JOURNAL Submitted (25-JUL-1994) Michael McClelland, California Institute of Biological Research, 11099 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES  
Source Location/Qualifiers

1..1489  
/organism="Leptospira weilii"  
/strain="serogroup hebdomadis, subserogroup Borlincana, serovar worsfoldi, strain Worsfold"  
/isolate="WAS2"  
/db\_xref="taxon:28184"  
<1..1489  
rRNA

BASE COUNT 383 a 332 c 455 g 316 t 3 others  
ORIGIN

Query Match 85.6%; Score 1263.8; DB 29; Length 1489;  
Best Local Similarity 94.1%; Pred. No. 1.7e-13;  
Matches 1376; Conservative 0; Mismatches 80; Indels 6; Gaps 6;

QY 1 gatcgtgctcagaactaacgctggcgcgcgctttaaacaatgcgaagtcgagcgggtag 60  
Db 16 GATCCTGGCTCAGAACTAACGCTGGCGCGCTCTTAAACATGCAAGTCAAGCGGAGTAG 75  
QY 61 caataccttagcggcgaacggtgagtaacacgt-ggtaattcttcctcagactcgtggata 119  
Db 76 CAATACTACGCGCGCAACGGGTGAGTAACACGTGGGTATCTTCCCGAGTCTGGGATA 135  
QY 120 acttcccaaaagaagctaataaccgtagtccctggtggatcacagaatttgatagata 179  
Db 136 ACTTTCCGAAAGGGAGCTAATACGTGGATGGTCCCGAGAGGTATATATTTTCGGGTA 195  
QY 180 aagattattgcttgagatgagccgcgcgcgattagctagttggtgaggttaattggctc 239  
Db 196 AAGATTATTGCTCGAGCTGAGCCGCGCGCGGATTAGCTAGTTGGTGAAGTAATGGCTC 255  
QY 240 accaaggcagatcggtagccgcgcctgagaggggtgctcggccacaa1ggaaactgagaca 299  
Db 256 ACCAAGGCGACATCGTAGCGCGGCTGAGAGGGTGTTCGGCCACANTGGAACCTGAGACA 315  
QY 300 cgttcatactctcagggagcagcagctgaagaattctgctcaatgggggaaacccctga 359  
Db 316 CGGTCCATACT-CTACGGGAGCAGCAGTTAGATATCTGCTCAATGGGGGAAACCCCTGA 374  
QY 360 agcagcgcgcgcgtgaacgaaggtcttcggattgttaagtctattaggca-ggaa 418  
Db 375 AGCAGCGACGCGCGTGAACGATGAAGGTCTTCGGATTGTAAAGTTCAATAAGCAGGGAA 434  
QY 419 aaataagcagaatgtgatgtgacgtgacctgacctaaagcaccggtacactacgtgccagca 478  
Db 435 AAATAAGCAGCGATGTGATGTGATGCTTACCTTGCCTAAAGCACCGGCTAACTTACCTGCCAGCA 494  
QY 479 gccgcggttaatacgtatggtgcaagcgttcttcggaatacttggtggtgctgaaagggtgcga 538  
Db 495 GCCGCGGTAAATACGTATGCTGTCAGCGTGTTCGGAATATTCAGGTGAAAGGTGCGGTA 554  
QY 539 gccgatttgaatcaggtgtaaaactgcgggtcgaacccgtggcctgcaacttgaac 598  
Db 555 GCGCGACATGTAAAGTCAGGTGTGAACACTGCGGGCTCAACTCGCAGCTGCACCTTGAAC 614  
QY 599 tacaagctggaatttggagaggaagtggaattccagggtgtagcgtgaaatcgtag 658  
Db 615 TATGTCTCTGGAGTTTGGAGAGGAAGTGGAAATTCACAGGTGAGCGGTGAATCGGTAG 674  
QY 659 atatctgaggaacaccagtgccgaagcagcttactggtcctaaactgaactgagaca 718  
Db 675 ATATCTGGAGAACACCACTGGCGAGGCGACTTCTGCGCCCTAAACACTGACCTGAGGCA 734  
QY 719 cgaagcgtgggttagtaaacgggttagatatacccggttaataccagccctaaactgtgctc 778  
Db 735 CGAAGCGTGGGTAGTGAACGGGATAGATACCCGGTAAATCCAGCCCTAAACGTTGCTC 794

QY 779 taccagttgttgggggtttttaaccctcagtaacgaacccaacccatgaattagtagaccgct 838  
Db 795 TACCAGTTGTTGGGGTTTTAAACCTTCAGTAACGAA-CTAACGGATTAAAGTAGACCGCT 853  
QY 839 ggggactatgctcgaagtgaaactcaaggaattgaagggttcgcaagaagggtg 898  
Db 854 GGGGACTATGCTCGCAAGAGTGAACCTCAAGGAATTGACGGGGGTNNNCACAAGCGGTG 913  
QY 899 gagcgtggttgaattcagatgatacccccaaaacctcaacctgggcttgacatgatctg 958  
Db 914 GAGCATGTGGTTTAAATTCGATGATACGGCAAAACCTCACCTTAGCTTGACATGGAGTGG 973  
QY 959 aatcatgttagagatatagagccttcgagcatatcagagtgctgcatggtgtcgta 1018  
Db 974 AATTATGTAGAGATACATCAG-CTTCGGGCCCTTTCACAGGTGCTGCATGTTGCTGCA 1032  
QY 1019 gctcgtgctgtagatgttgggttaagtcccccaacgaagcgaacccctatctgtatgtg 1078  
Db 1033 GCTCCTGCTGAGATGTGGGTTAAGTCCCGACAGAGCGCAACCTCACCTTATGTTG 1092  
QY 1079 cta-ctttaagttggcactggtacgaaactcccggtagaaccggaggaaggcgga 1137  
Db 1093 CCATCATTTAGTTGGGCACTCGTAAGGAACCTGCCGTGACAAACCGGAGGAGCGGA 1152  
QY 1138 tgaactcaaaatcctcatggcctttatgtccaggggccacacacgctactacaa1ggccgata 1197  
Db 1153 TGACGTCAAACTCTCATGGCCTTTATGTCTAGGCGCAACACAGCTGCTACAAATGGCGGTA 1212  
QY 1198 cagagggctgcgaactcgaagagggagctaactctctaaagtcggtcccagttcgatt 1257  
Db 1213 CAAGGGTAGCCAACTCGGAGGGGAGCTAATCTCAAAAACCGGTCCTCCAGTTCGGATT 1272  
QY 1258 ggggtctgcaactcgaaccccatgaagtcggaatcgtagtaatcgcggatcgatcgccg 1317  
Db 1273 GGAGTCTGCAACTCGACTCCATGGAAGTGGGAATGCTAGTAATCGCGGATAGCATGCCG 1332  
QY 1318 cgtgaatactgctccggaccttgtacacacgcgcgcgtcacacccacctgagtgggagaca 1377  
Db 1333 CGGTGAATACGTTCCCGGACCTTGTACACACCGCGCTCACACACCTGAGTGGGAGCA 1392  
QY 1378 cccgaagtggtcttcttgaacggaagagacagactactaaggtgaaactcgtaaagg 1437  
Db 1393 CCCGAAGTGGTCTTTGCCNACCGTAAGGAGCAGACTACTAAGGTGAAACTCGTGAAGGG 1452  
QY 1438 ggtgaagtgtgaacaaggtaac 1459  
Db 1453 GGTGAAGTCGTAAACAAGGTAGC 1474

RESULT 7

LBUL2670  
LOCUS 1513 bp DNA BCT 18-FEB-1995  
DEFINITION Leptospira borgpetersenii Hardjo bovis/Sponselee 16S rRNA gene, partial sequence.  
ACCESSION U12670  
VERSION U12670.1 GI:558929  
KEYWORDS  
SOURCE Leptospira borgpetersenii.  
ORGANISM Leptospira borgpetersenii  
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.  
REFERENCE 1 (bases 1 to 1513)  
AUTHORS Ralph.D. and McClelland,M.  
TITLE Phylogenetic evidence for horizontal transfer of an intervening sequence between species in a spirochete genus  
JOURNAL J. Bacteriol. (1994) In press  
AUTHORS 2 (bases 1 to 1513)  
TITLE Direct Submission  
JOURNAL McClelland,M.  
AUTHORS  
TITLE Submitted (25-JUL-1994) Michael McClelland, California Institute of Biological Research, 11099 North Torrey Pines Road, La Jolla, CA 92037, USA  
FEATURES  
source Location/Qualifiers  
1..1513

/organism="Leptospira borgpetersenii"  
/strain="serovar hardjo, Strain Hardjo bovis/Sponselee"  
/isolate="BD50"  
/db\_xref="taxon:174"  
<1..1513  
/product="16S rRNA"  
BASE COUNT 385 a 346 c 315 t 3 others  
ORIGIN

Query Match 85.5%; Score 1262.2; DB 2: Length 1513;  
Best Local Similarity 94.0%; Pred. No. 1.8e-13;  
Matches 1375; Conservative 0; Mismatches 81; Indels 6; Gaps 6;  
Qy 1 gatcatggctcagaactaaagctggcgccgctcttaaacatcaagtcagcgggtag 60  
Db 40 GATCCTGGCTCAGAACTAACCTGGCGCGCTCTTAACATGCAAGTCAGCGGAGTAG 99  
Qy 61 caatcacctagcggcgaacgggtgagtaacacgt-ggtaactctctcccgagctcgggata 119  
Db 100 CAATACTCAGCGGCGGAGCGGTGAGTACAGCTGGGTAACTCTCCCGAGCTCGGGATA 159  
Qy 120 actttccgaaggaagactaataccgagtagtctctgttggatcacaaagattttaggta 179  
Db 160 ACTTTCCGAAGGGGAGCTAATCTGATAGTCCCGAGAGCTATAGATTTTTCGGGTA 219  
Qy 180 aagatttattcttgagatgagcccgccgagcttagctagttggtgagtaagtgc 239  
Db 220 AAGATTTATGCTGGCGCTGAGCCCGCGCGGATAGTACTGTTGGTGGAGTAAATGGCTC 279  
Qy 240 accaagggcagcagtcggtagcggcctgagaggggtgtccggcccaactggaactgagaca 299  
Db 280 ACCAAGGCGAGCATCGGTAGCGGCTCAGAGGGTGTTCGGCCCAATGGAAGTGAAGACA 339  
Qy 300 cggctcactctcagggagggcagcagtaagaatcttgctcaatggggaaacctga 359  
Db 340 CGGTCCATACT-CTACGGGAGCGACAGCTTAAGAACTTCGTAATGGGGGAAACCTTGA 398  
Qy 360 agcagcgacgcgcgtgaacgaagagctctcgattgtaaaagtccattaggca-ggaa 418  
Db 399 ACCAGGAGCGCGCGTGAACGATGAAGCTTCGGGATTGTAAGTTCAATAAGCAGGGAA 458  
Qy 419 aaataagcagcaatgtgatgtgtaactgtcctaagcaccggctaaactagtcgccaaga 478  
Db 459 AAATAAGCAGCGATGTGATGTGATGTGACCTGAAGCACCAGCTTAACCTAGTCCGACGA 518  
Qy 479 gccgcgtaatacgtatgtgcaacgctgttcggaatcattggcgtaaaaggggtgcga 538  
Db 519 GCGCGGTAATACGTATGTGCAACGCTTTCGGAATCATTTGGCGCTAAAGGTTGCGTA 578  
Qy 539 ggcggatttgaagtcaggtgtaaaactgcgggctcaacccgtggcctgcactgaaac 598  
Db 579 GCGGACATGTAACTCAGTGTGAAACTGCGGGCTCAACTCGCAGCTGCACCTTGAAC 638  
Qy 599 tacaagctcggagtttggagaggaagtggaatccagggttagcgggtgaaatcgtag 658  
Db 639 TATGTCTCGGAGTTTGGGAGAGGAAGTGGAAATTCAGGTGTAGCGGTGAATCGGTAG 698  
Qy 659 atatctggaggaacaccagtcgcaagggcagcttgcgtcctaaacagcgtgagga 718  
Db 699 ATATCTGGAGGAACACCACTGCGGAGGCGCACTTCTGGCCTAAACTGACCTGAGGCA 758  
Qy 719 cgaagcgtgggttagtaaacgggattagatacccccgggtaattccacgcccctaacscttgc 778  
Db 759 CGAAGCGTGGGTAGTGAACGGGATTAGATACCCGGTAATCCAGCCCTAAACCTTGTCT 818  
Qy 779 taccagtttgggggttttaacctcagtaacgaacctaacggattagtagcgcct 838  
Db 819 TACCAGTTGTGGGGGTTTTTAACTCTCAGTACGAA-CTAACGGATTAAGTAGACCGCT 877  
Qy 839 ggggactatgctcgaagagtgaaactcaaaaggaattgacgggggtccgcacaaagcgtg 898  
Db 878 GGGGACTATGCTCGCAAGAGTGAAGTCAAGGAAATTCAGCGGGGTNNNCAACAAAGCGGTG 937

Qy 899 gagcatgtgttttaattcagatgatacccccaaaacccacctcggcgttgacatggtctg 958  
Db 938 GAGCATGTGTTTAACTTCATGATAGCGGAAACCACTACCTAGCTTGACATGGAGTGG 997  
Qy 959 aatcatgtagagatatagagccttcggcgagattcaacagggtgctgcatggttgcgtca 1018  
Db 998 AATCATGTAGAGATACATAGAG-CTTCGGGCCGCTTTCACAGGTCCTGCATGGTGTGCTCA 1056  
Qy 1019 gctcgtgctgagatgttgggttaagtcccccaagcagcgcaacccctatcgatgttg 1078  
Db 1057 GCTCGTGTCTGAGATGTTGGTTAACTCCCGCAACGAGCGCAACCTCACCTTATGTTG 1116  
Qy 1079 cta-cttaagttgggacactggtacgaaactgcgggtgacaaacccgaggaaggggga 1137  
Db 1117 CCATCATTCAGTTGGGCACTCGTAAAGAACTGCCGTGACAAACCGAGGAGCGGGGA 1176  
Qy 1138 tgcgtcaaatcctcagctgcttattatgcaggggcacacagctgctacaatggccgata 1197  
Db 1177 TGAGCTCAAACTCCTATGGCCCTTATGCTAGGGCAACACACGTCTACAAATGGCCGGTA 1236  
Qy 1198 cagaggtgcgccaactgcgaagggaggtcaatctctaaaagtcggtcccagttcgatt 1257  
Db 1237 CAAAGAGTAGCCCAACTCGCGAGGGGAGGTAACTCAAAAGCGCGTCCAGTTCGGATT 1296  
Qy 1258 ggggtctgcaactcgacccccctgaagtgcggaaatcgctagtaatcgcggtatcgatgc 1317  
Db 1297 GGAGTCTGCAACTCGACTCCATGAAGTCGGAAATCGCTAGTAATCGCGGATCAGCATGCCG 1356  
Qy 1318 cgggtgaatagttcccgacacttgtaacaccccgctcacacacactagtgaggagca 1377  
Db 1357 CGGTGAATAGTTCCTCCGACCTTGTACACACCCGCTCACACCTAGTGGGAGCA 1416  
Qy 1378 ccgaaagtgccttctgttaaccgttaagagagacagactactaaagtgaaactcgtaaagg 1437  
Db 1417 CCCGAAGTGTCTTTGGCAACCCGCAAGGAAGCAGACTACTAAGTGAAGTCTGTGAAGG 1476  
Qy 1438 ggtgaagctgtaacaaggatcc 1459  
Db 1477 GGTGAAGTCGTAAACAAGGTAGC 1498

## RESULT 8

LBUI2669 18-FEB-1995  
LOCUS  
DEFINITION Leptospira borgpetersenii 1627 Burgas 16S rRNA gene, partial  
sequence.  
ACCESSION U12669  
VERSION U12669.1 GI:558928  
KEYWORDS  
SOURCE Leptospira borgpetersenii.  
ORGANISM Leptospira borgpetersenii  
Bacteria; Spirochaetales; Leptospiroaceae; Leptospira.  
REFERENCE 1 (bases 1 to 1489)  
Ralph, D. and McClelland, M.  
Phylogenetic evidence for horizontal transfer of an intervening  
sequence between species in a Spirochete genus  
J. Bacteriol. 176 (19), 5982-5987 (1994)  
MEDLINE 95014031  
AUTHORS McClelland, M.  
REFERENCE 2 (bases 1 to 1489)  
Direct Submission  
JOURNAL Submitted (25-JUL-1994) Michael McClelland, California Institute of  
Biological Research, 11099 North Torrey Pines Road, La Jolla, CA  
92037, USA  
FEATURES  
source  
1. 1489  
/organism="Leptospira borgpetersenii"  
/strain="serovar balcanica, strain 1627 Burgas"  
/isolate="BD30"  
/db\_xref="taxon:174"  
<1..1489  
/product="16S rRNA"  
rRNA



BASE COUNT 385 a 330 c 454 g 311 t 9 others  
ORIGIN

Query Match 85.4%; Score 1261; DB 2: Length 1489;  
Best Local Similarity 93.8%; Pred. No. 1.9e-13;  
Matches 1372; Conservative 0; Mismatches 84; Indels 6; Gaps 6;

Qy 1 gatcgtgctcagaactaacgctgctggcgogtcttaaacatgcaagtcgagcggggtag 60  
Db 16 GATCCTGGCTCAGAACTAACGCTGGCGCGGCTCTTAAACATGCAAGTCAAGCGGAGTAG 75

Qy 61 caatacctagcgcgaacggtgagtaacagct-ggtaattctctccagctctgggata 119  
Db 76 CAATACTCAGCGGGCAACGGGTGAGTACACGTGGGTATCTCTCCAGTCTGGGATA 135

Qy 120 actttccgaagaagcgttaacccgagatagtcctgtgtggatcacaaagtattgaggtta 179  
Db 136 ACTTTCCGAAGGGGAGCTAATACTTGGATAGTCCCGAGAGGTCATAGGATTTTTCGGGTA 195

Qy 180 aagatttattgcttgagatgagcccgcggtgatttagctagttggtgaggttaattggttc 239  
Db 196 AAGATTATTTGCTCGGAGATGAGCCCGCGCGGATTAGCTAGTTGGTGAAGGTAAATGGCTC 255

Qy 240 accaagcgagcagatcggtagccgcctagaggggtgctcgcccaaatggaactgaaga 299  
Db 256 ACCAAGCGACATCGGTAGCCCGGCTGAGAGGGTGTTCGGCCACANTGGAAGTGAAGTA 315

Qy 300 cggctcactactcctacggagcagcagtttaagaatcttgcctcaatg9gggaaacccctga 359  
Db 316 CGGTCCATACT-CTACGGGAGGACAGCTTAAGATCTTGCTCAATGGGGGAAACCTGA 374

Qy 360 agcagcgacgcgctgtaacgaagaagttcttcgagatgttaaaagtctattagggca-gaaa 418  
Db 375 AGCAGCGACGCGCGCTGAACGATGAAGGTCTTCGGATTGTAAAGTTCAATAAGCAGGAA 434

Qy 419 aaataagcagaatgtagtgactgctcctaagcagcagcgaactacgtgccaaga 478  
Db 435 AAATAAGCAGCATGTGATGATGGTACCTGCTTAAGCACCAGGCTAACTACGTGCGCAGCA 494

Qy 479 gccgcggttaatacgtatggtgcaagcgtgttcggaatcattggcgtaaaaggtgcgta 538  
Db 495 GCCGCGGTAAATACGTATGTTGCGAAGCGTGTTCGGAATCATTTGGCGTTAAAGGGTCCGTA 554

Qy 539 ggcggattgttaagtcaggtgtgaaactgogggctcaaccogtgcctgcacttgaac 598  
Db 555 GCGCGACATGTAAATGACGTGTGAAACTGCGGGCTCAACTCGCAGCGCTGCACTTTGAAC 614

Qy 599 tacaagctgagtttggagagcaagtggaattccaggtgtagcgtgaaatcgtag 658  
Db 615 TATATGCTGGAGTTTGGAGAGGCAAGTGGAAATTCAGGTGACGGTGAATGCGGTAG 674

Qy 659 atatctgagggaacaccagtgccgagcagcttgcctgagcctaaactgacgtgagga 718  
Db 675 ATATCTGGAGGAACACCACTGCGGAAGCGGACTTGTGCGCTTAAACTGACGCTGAGGCA 734

Qy 719 cgaagcgttggtagtaaacgggataagatacccggttaattccagccctaaactgttgc 778  
Db 735 CGAAAGCGTGGGTAGTAACGGGATTAGATACCCCGGTAATCCAGCGCTAAACGCTTGTCT 794

Qy 779 taccagttgtgggggttttaaccctcagtaacggaactaacggttaagttagcgcct 838  
Db 795 TACCAAGTTGTGGGGTTTTTAACTCTAGTAAACGAA-CTAACGGATTAAAGTAGACCGCT 853

Qy 839 ggggactatgctcgaagtagtaaaactcaaggaattgacgggggtccgcacaaagcgggtg 898  
Db 854 GGGGACTATGCTCGCAAGAGTCAAACTCAAAAGGAATTGACGGGGTNNNCAACACGGGTG 913

Qy 899 gagcatgtggttttaattcgatgatacccccaaaactcaactcctggggttgacatggatctg 958  
Db 914 GAGCATGTGTTTAAATTCGATATACGGCAAAACCTCACCTAGGCTTGACATGGAGTGG 973

Qy 959 aatcatgtagaagatatagagccttcgggcagagattcaacaggtgctgcatggtgtgcgta 1018

Db 974 AATCATGTAGAGATACATGAG-CTTTCGGGCCCTTCACAGGTGCTGCATGGTGTGCTCA 1032  
Qy 1019 gctcgtgctgagatgttgggttaagtcgcgcaacgagcgcaaccctctatcgatgttg 1078  
Db 1033 CCTCGTGTCTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCTTCACCTTATGTTG 1092  
Qy 1079 cta-ccctaagttggcactggtacgaaactgcccgggtgacaaaaccggagggcggga 1137  
Db 1093 CCATCATTCAGTTGGGCACTCGTAAGGAACCTGCCGTGACAACTGGAGAAAGCGGGA 1152  
Qy 1138 tgacgtcaaatcctcatgacctttatgtccagggccacacacgtgctacaatggccgata 1197  
Db 1153 TGACGTCAANNNNNATGCCCTTATGTCTAGGGCAACACAGCTGCTACATGGCCGGTA 1212  
Qy 1198 cagagggctgcgcaactcgaagaggaggttaactctctaaaagtcggtcccagttcggatt 1257  
Db 1213 CAAAGGGTAGCAACTCGCAGGGGAGCTAATCTCAAAAAGCCGCTCCCACTTCGGATT 1272  
Qy 1258 ggggtctcgaactcgaacccatgaagtcggaatcgtcgtagtaatcgcgatacgcgcg 1317  
Db 1273 GGAGTCTCAACTCGACTCCATGAATCGGAATCGTATAGTAATCGCGATCAGCATGCG 1332  
Qy 1318 cgtgaatacgttcccgcgacctgtacacacgcgcgtcacaccacctgagtgaggaga 1377  
Db 1333 CGGTGAATACGTTCCCGGACCTTTGTACACACCGCCGCTCACACCACCTGAGTGGGAGCA 1392  
Qy 1378 cccgaagtggtgtttgttaaccgttaagagagacagactactaaggtgaaactcgtaaaagg 1437  
Db 1393 CCCGAAGTGGTCTTTCCCAACCGCAAGGAAGCAGACTACTAAGGTGAAACTCGTGAAGG 1452

Qy 1438 ggtgaagtcgttaacaaggtacc 1459  
Db 1453 GGTGAAGTCGTAAACAAGGTAGC 1474

RESULT 9  
LNUI2671  
LOCUS LNUI2671 1516 bp DNA BCT 18-FEB-1995  
DEFINITION Leptospira noguchii Fort Bragg 16S rRNA gene, partial sequence.  
ACCESSION U12671  
VERSION U12671.1 GI:558930  
KEYWORDS  
SOURCE Leptospira noguchii.  
ORGANISM Leptospira noguchii  
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.  
REFERENCE 1 (bases 1 to 1516)  
Ralph,D. and McClelland,M.  
AUTHORS Phylogenetic evidence for horizontal transfer of an intervening  
sequence between species in a spirochete genus  
J. Bacteriol. 176 (19), 5982-5987 (1994)  
TITLE  
JOURNAL  
MEDLINE 95014031  
REFERENCE 2 (bases 1 to 1516)  
McClelland,M.  
AUTHORS Direct Submission  
TITLE  
JOURNAL Submitted (25-JUL-1994) Michael McClelland, California Institute of  
Biological Research, 11099 North Torrey Pines Road, La Jolla, CA  
92037, USA  
FEATURES  
Location/Qualifiers  
1..1516  
/organism="Leptospira noguchii"  
/strain="serogroup Autumalis, subserogroup Fort Bragg,  
serovar fortbragg, strain Fort Bragg"  
/isolate="NB36"  
/db\_xref="taxon:28182"  
<1..1516  
/product="16S rRNA"  
BASE COUNT 392 a 342 c 462 g 314 t 6 others  
ORIGIN

Query Match 85.4%; Score 1260.8; DB 2: Length 1516;  
Best Local Similarity 93.9%; Pred. No. 1.9e-13;

Matches 1373; Conservative 0; Mismatches 83; Indels 6; Gaps 6;

```
Qy 1 gatcatgctcagaactaacgctgctgctgctcttaaacatgcaagtcgagcgggtag 60
Db 43 GATCTGGCTCAGAACTAACGCTGCGCGCTCTTAACATGCAAGTCAAGCGAGTAG 102
Qy 61 caatcacctagcgcgaacggtgagtaacacgt -ggtaatctctccagctctgggata 119
Db 103 CAATACTCAGCGCGCAACGGGTGAGTACACGTGGGTAACTCTTCNCGAGTCTGGGATA 162
Qy 120 actttccgaagaagaagctaaaccggtagctctgttggatcacagaatttgatagta 179
Db 163 ACTTTCCGAAGGGAAGCTAATAGTGGTAGTCCCGAGAGATCAATAGATTTCCTGGGATA 222
Qy 180 aagattattgttggagatgagcccgccgctgattagctagttggtgaggttaattggctc 239
Db 223 AAGATTTCATTGCTCGAGATGAGCCCGCTCCGATTAGTTGTTAAGGTAAGGCTT 282
Qy 240 accaagcgacgacgctgtagccggcctgagagggctgtagcccaaatggaactgagaca 299
Db 293 ACCAAGCGACGATCGGTAGCGCGCTCAGAGGGTGTTCGGCCACAATCGAACTGAGACA 342
Qy 300 cggctccactactcctcgggagcgagcagcttaagaatcttctcaatgggggaaaccctga 359
Db 343 CGGTCCATACT-CTACGGGAGCGACACTTAAGATCTTGCTCAATGGGGGAACCCCTGA 401
Qy 360 agcagcgacgcgcgcgtgaacgaagaagctctcggattgtaaaagttcattaggca -ggaa 418
Db 402 AGCAGCGACGGCGGTGAACGATGAAGTCTTCGGATTGTTAAAGTTCAATAGCAGGAA 461
Qy 419 aataagcagcaatgtgatgtgtaacctgctaaagcgcgctaaactacgtgcccagca 478
Db 462 AAATAAGCAGCAATGTGATGATGTACCTGCTTAAGCACCGGCTAACTAGTCGCCAGCA 521
Qy 479 gccgggttaatacgtatgttgcgaagcgttgttcggaatcattggcgctaaaggggtgcgta 538
Db 522 GCCGGGTAAATACGTATGTGTCGAAGCGTGTTCGGAATCATTTGGCGGTAAAGGGTGCCTA 581
Qy 539 ggcggatttgaagtcaggtgtaaaactcggcgtcaaccgctgacctgcaactgaaac 598
Db 582 CGCGGACATGTAAGTCAGCTGTGAACCTGGGGGCTCAACTCCCGGCTGCACTTGAAC 641
Qy 599 tacaagctgagttggagagcgaagtgaattccagggtgtagcgtgaaatgcgtag 658
Db 642 TATCTGTCTGAGTTTGGGACAGCAAGTGAATTCACAGTGTAGCGGTGAATGCGTAG 701
Qy 659 atactggaggaacacacgctggcgaagcgactgtgctgctcctcaaaactgacgtgagca 718
Db 702 ATATCTGGAGGAACACCACTGCGGAAAGCGACTTGTCTGGCTAAACTGACGCTGAGGCA 761
Qy 719 cgaagcgtggtgtagtaaacgggtagataccgcgtaattccacgcgcctaaacgttgc 778
Db 762 CGAAGCGTGGGTAGTGAACGGGATAGATNCCCGGTAACTCCAGCCCTAAACGTTGTC 821
Qy 779 taccagttgttgggggttttaacctcagtaacgaaactaacgattaaagtagaccgct 838
Db 822 TACCAGTTGTTGGGGGTTTTAACCTCAGTAACGAA-CTAACGGATTAAAGTAGACCGCT 880
Qy 839 gggactatctcgaagagtgaaactcaaaaggaattgacgggggtccgcacaagcggtg 898
Db 881 GGGACTATGCTCGCAAGTGAAGTCAAAAGGAATTGACGGGGGTNNNCACAAAGCGGTG 940
Qy 899 gagcatgtggttgaattcgaatgatacccccacaaacctcaactgggttgacatggtatcg 958
Db 941 GAGCATGTGGTTTAAATTCGATGATACCGGAAACCTCACCTAGGCTTGACATGAGTGG 1000
Qy 959 aatcatgtagagatatatgagctctcggcagattcagagtgctgcatggttgcgtca 1018
Db 1001 AATCATGTAGAGATACATAGAG-CTTCGGGGCGCTTCACAGGTGCTGTCATGTTCTCGTCA 1059
Qy 1019 gctcgtcgtgagatgttgggttaagttcccgcaacgagcgcaacctatcgtatgtg 1078
Db 1060 GCTCGTGTCTGATGATGTTGGGTTAAGTTCGGCAACGAGCGCAACCTCACCTTATGTNG 1119
```

```
Qy 1079 cta-ccctaaagttggcactggtacgaaactgctggttgcacaaacggaggaagcgggga 1137
Db 1120 CCATCATTCAGTTGGCACTCGTAAGGAACATGCCGCTGCACAAACGGAGGAGCGGGGA 1179
Qy 1138 tgaagtcataactcctcatggtccctttatgtccaggggcccacacgctgctacaatggccgata 1197
Db 1180 TGAGTCAAACTCCTCATGGCCTTTATGTCTAGGGCAACACACAGCTGCTACANTGGCCGTA 1239
Qy 1198 cagagggctcgccaactcgcaagaggggaggttaattctctaaaaagtcgggtcccgattcggtatt 1257
Db 1240 CAAAGGGTAGCCAACCTCCGAGGGGAGCTAATCTCAAAAAGCGGTCCCAAGTTCGGATT 1299
Qy 1258 ggggtctgcaactcgaccccatgaagtcggaatcgtagtaactcgcgagatcagcatgccg 1317
Db 1300 GGAGTCTCAACTCGACTCCATGAAGTCGGANTCGTAGTAACTCCGGATCAGCATGCCG 1359
Qy 1318 cgggtgaatacgttccccgagcctgtacacaccgcccgtcacaccacctgagtggggagca 1377
Db 1360 CGGTGAATACCTTCGGGACCTTGTACACACCGCGCTCACACCACCTGAGTGGGAGCA 1419
Qy 1378 cccgaagtggtcttgttaaccgtaaggaagagacagactactaagtgaaactcgttaagggg 1437
Db 1420 CCCGAAGTGGTCTTGGCCACCGCAAGGAGCAGACTACTAAGGTGAACCTCGTGAAGG 1479
Qy 1438 ggtgaagtcgtacaaggtacc 1459
Db 1480 GGTGAAGTCGTAAACAAGGTAGC 1501
```

## RESULT 10

## LB16SRDNP

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## rRNA

## BASE COUNT

## ORIGIN

## Query Match

## Best Local Similarity

## Matches 1343; Conservative

## 84.9%; Score 1254.6; DB 2; Length 1425;

## Pred. NO. 2.5e-13;

## Mismatches 76; Indels 4; Gaps 4;

## Qy 16 ctaacgctggcgctgcttaaacatgcaagtc -gagcgggtagcaatcacctagcgcc 74

## Db 1 STARCGTGGCGCGCTCTTAACATCCCAATCCAAAGCGGAGTAGCAATACTACGCGGC 60

## Qy 75 gaacgggtgagtaaacagct -ggtaatctctccagtcgagtcgtggataacttccgaagga 133

## Db 61 GAACGGGTGAGTAACAGCTGGGTAACTTCTCCGAGTCTGGGATAACTTTCGGAAGGG 120

LB16SRDNP 1425 bp DNA BCT 02-DEC-1993  
L.borgpetersenii gene for 16S ribosomal RNA (partial).  
221630  
221630.1 GI:433579  
16S ribosomal RNA.  
Leptospira borgpetersenii.  
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.

REFERENCE 1 (bases 1 to 1425)  
Hooke, J.V.

Phylogeny of Leptospiraceae and related Spirochaetes

Unpublished

2 (bases 1 to 1425)

Hooke, J.V.

Direct Submission

Submitted (09-FEB-1993); Hooke J.V., Public Health Laboratory

Service, Leptospira Reference Laboratory, Stonebow Road, Hereford,

Herefordshire, United Kingdom, HR1 2ER

Location/Qualifiers

1. 1425

/organism="Leptospira borgpetersenii"

/strain="JAVANICA VELDRAT BATAVIA 46"

/db\_xref="taxon:174"

<1..>1425

/product="16S ribosomal RNA"

369 a 327 c 430 g 297 t 2 others



QY 134 aagctaataaccggatagctctgttgatgatacaagaatttgataggttaagattattatgttt 193  
DB 121 GAGCTTAATACTGGATAGTCCCGAGAGGTATAGGATTTTTCGGGTAAGAAATTTATGTCTC 180  
QY 194 ggaatgaagcccgccgagattagctagctgaggttaagctcaccacagcgacgat 253  
DB 181 GGAGATGAGCCCGCCCGGATTTAGCTAGTTGTTGAGGTAATGGCTTACCACCGCCACGAT 240  
QY 254 cgttagcggcctgagaggggttcggccacacaaatggaactgagacaggtcactact 313  
DB 241 CGGTAGCCGGCTGAGAGGGTCTTCGGCCACAAATGGAACGTGAGACAGCGTCCATCTCT 300  
QY 314 acggaagcagcagttgaagaattctgctcaatgggggaacccctgaagcagcagcgccg 373  
DB 301 ACGGAGGCGAGCTTAAAGATCTTGCTCAATGGGGGAACCCCTGAAGCAGCGAGCCGCG 360  
QY 374 gtaacgaagaaggtcttcggattgataagttcattagga-ggaataaagcagcaat 432  
DB 361 GTGAACGATGAAGTCTTCGGATTGTAAAGTTCAATAAGCAGCGGAATAATAGCAGCGAT 420  
QY 433 gtgatgttggtacctgcctaaagcaccggcctaactacgtgccagcagccgcggttaacg 492  
DB 421 GTGATGATGTTACCTGCCTAAAGCAGCCGCTAATCTAGCTGCCAGCAGCGCGGTAAACG 480  
QY 493 tatgtgcaagcgttgttcggaaatcattggcgtaaaaggtgctgagcggtttgtgaag 552  
DB 481 TATGTTGCAAGCGTTGTTGCGAATCATTTGGCGTAAAGGGTTCGTAGGCGACATGTAG 540  
QY 553 tcagttgtgaacactgcggctcaaccgtggcctgcactgaaactcaagctcagagt 612  
DB 541 TCAGTGTGAACACTCGCGGCTCAACTCCAGCCCTGCACCTTGAACATATGTCCTGAGGT 600  
QY 613 ttggagagcgaagtggaaattccaggtgtagcgtgaaatgcgtagatatctgggggaac 672  
DB 601 TTGGGAGAGCGAAGTGAATTCACGGTGTAGCGGTGAATGCTAGATATCTGGAGGAAC 660  
QY 673 accagtggcgaagcgactgtgtgctcaaaactgacgtgagggcagcaaaagcgtgggta 732  
DB 661 ACCAGTGGCGAAGGCGACTGTGCGCTAAACTGACGCTGAGGACGCAAAAGCGTGGGTA 720  
QY 733 gtaacgggattagatacccggttaaccacgctcaacgcttctaccagttgtgtgg 792  
DB 721 GTGAACGGGATTAGATACCCCGGTAACTCACGCCCTAAACGTTGTCTACCACTTGTTCGG 780  
QY 793 ggtttaaccctcagtaacgaacctaaacggttaagtagagccgctggggaactatgctcg 852  
DB 781 GGTTTTAAACCTCAGTACGAACCTAACGGATTAAGTAGACCGCCTGGGGACTATGCTCG 840  
QY 853 caagagtaaaactcaagaaattgacggggtccgcacaaagcggtagacatgtgttta 912  
DB 841 CAAGAGTGAACCTCAAGGAATTGACGGGGGTCCGCACAAAGCGGTGGAGCATGTGTTTA 900  
QY 913 attcagatatacccaaaacctcactcgttgacttgacatggatcgtgaatcatgtagat 972  
DB 901 ATTCGATGATACCGAAGAACCTCACCATTAGGCTTGACATGGAGTGAATCATGTAGAGAT 960  
QY 973 atatagccttcgggcagatcacaaggtgctgcatgggttgctgacgtcgtgctgtag 1032  
DB 961 ACATGAGCCTTCGGGCGGCTTACAGAGTGTGCTGATGTTGTGCTGAGCTGCTGTGAG 1020  
QY 1033 atgttggttaagttccgcacagcagcgaacccctatgctatgttgcta-ccctaaagttg 1091  
DB 1021 ATGTTGGGTTAAGTCCCGCAACGAGCGCAACCTCACCCTATGTTGCCATCATCAGTTG 1080  
QY 1092 ggcactggttacgaactgcggtgacaaacggaggaagcgggatgacgtcacaatcct 1151  
DB 1081 GGCACCTGTAAGGAACCTGCCGTGACAAACCGAGGAAGCGGGATGACGTCAAACTCT 1140  
QY 1152 catggcctttatgtccagggccacacagctgctcaaatggccgaacacagaggggtcgccaa 1211  
DB 1141 CATGGCCTTTATGTCTAGGGCAACACACGCTGCTACAAATGGCCGGTACAAAGGGTAGCCAA 1200  
QY 1212 ctgcgaagaggagtaactctctaaagtcggtccagttcggtatgggtctgcaactc 1271

DB 1201 CTCGGCAGGGGAGGTAAATCTCAAAAGCCGTCCTCAGTTCCGATTTGGAGTCTGCAACTC 1260  
QY 1272 gacccataagtcggaatcgtagtaatcgcggtacagcatgcgagcgcgtgaatcagcttc 1331  
DB 1261 GACTCCATGAAGTCGAATCGCTAGTAATTCGGGATCAGCATGCCGGGTGAATAGCTTC 1320  
QY 1332 cggacctgtgtacacacccgctcacacccactgagtgaggagcaccgcaagtgtctt 1391  
DB 1321 CGGACCTTGTACACACCCCGTACACACCTGAGTGGGAGCACCAGGAGTGTCTT 1380  
QY 1392 tattaacctgaaggagacagactactaagtgtaactcgtaaagg 1436  
DB 1381 TGCCAACCCCAAGGAGCAGACTACTAAGGTGAACCTCGTGAAGG 1425  
  
RESULT 11  
WB4616S  
LOCUS WB4616S 1486 bp DNA BCT 18-FEB-1995  
DEFINITION Leptosira weilii Sarmin 16S rRNA gene, partial sequence.  
ACCESSION U12673  
VERSION U12673.1 GI:558932  
KEYWORDS  
SOURCE Leptosira weilii.  
ORGANISM Leptosira weilii  
Bacteria; Spirochaetales; Leptospiraceae; Leptospira.  
REFERENCE 1 (bases 1 to 1486)  
AUTHORS Ralph,D. and McClelland,M.  
TITLE Phylogenetic evidence for horizontal transfer of an intervening  
sequence between species in a Spirochete genus  
J. Bacteriol. 176 (19), 5982-5987 (1994)  
JOURNAL 95014031  
MEDLINE  
REFERENCE 2 (bases 1 to 1486)  
AUTHORS McClelland,M.  
TITLE Direct Submission  
JOURNAL Submitted (25-JUL-1994) Michael McClelland, California Institute of  
Biological Research, 11099 North Torrey Pines Road, La Jolla, CA  
92037, USA  
  
FEATURES  
source  
1..1486  
/organism="Leptosira weilii"  
/strain="serogroup Icterohaemorrhagiae, subserogroup  
Sarmin, serovar sarmin, strain Sarmin"  
/isolate="WB46"  
/db\_xref="taxon:28184"  
<1..1486  
/product="16S rRNA"  
BASE COUNT 384 a 332 c 452 g 315 t 3 others  
ORIGIN  
  
Query Match 84.8%; Score 1251.8; DB 29; Length 1486;  
Best Local Similarity 94.0%; Pred. No. 2.7e-13;  
Matches 1375; Conservative 0; Mismatches 80; Indels 7; Gaps 7;  
  
QY 1 gatcatgctcgaactaacgctgcggcgctcttaaacatgcgaagtcgagcgggtag 60  
DB 14 GATCTGTGCTCAGACATAACGCTGGCGGCGCTCTTAAACATGCAAGTCAAGCGGATAG 73  
QY 61 caatacctagcgcggaacggtgagtaacagct-ggtaactcttcctccagctcggtata 119  
DB 74 CAATATCCACGCGGCAACGGGTGAGTAACACGCTGGGTAAATCTTCTCCGAGTCTGGGATA 133  
QY 120 acttcccgaaaggaagcctaaccggaatagctcctgttgatcacagaatttgataggtta 179  
DB 134 ACTTTCCGAAGGGGAGCTTAATCTGGATAGTCCCGAGAGGTCATATGATTTTCCGGTA 193  
QY 180 aagattattgtttgagatgagcccgccgagattagcttagttggtgaggttaagcttc 239  
DB 194 AAGATTTATTTGTCGAGCTGAGCGCCGCGCGCTAGTAGTTGTTGAGTAAATGGCTC 253  
QY 240 accaaggcgacatcggttagccggcctgaggggtgtccggccacaaatgaactgagaca 299

```
Db 254 ACCAAGCGCAGCATCGGTAGCCGCCCTGAGAGGGTGTTCGGCCACAATGAACTGAGACA 313
Qy 300 cggctccatactcctacgggagcagcaggttaagaattctgctaattggaagaaacctga 359
Db 314 CGGTCCATACT-CTACGGGAGCGAGCAGTCTAAGAACTCTTGCTCAATGGGGGAAACCCCTGA 372
Qy 360 agcagcagcgcgcgtgtaacgaagaaggtctctcgattgtaaaagtctcatlagga-ggaa 418
Db 373 AGCAGCGAGCGCGCTGAACCATCAAGGCTCTCGGATGTAAACTTCAATAAGCAGGGAA 432
Qy 419 aataagcagcaatgtgatgtatgtctgtaactgcctaagaacacgcggtaactaactgcgaagca 478
Db 433 AAATAAGCAGCAATGTGATGATGTACTGCTTAATCACCGGCTAACTAGTGCCAGCA 492
Qy 479 gcccggttaatacatgtgtcgaagctgtgtcgaaatcattggcgttaaggtgcgta 538
Db 493 GCCCGGTAAATACGTATGCTGCAAGCGTGTTCGGAATCATTTGGCGGTAAAGGTGGGTA 552
Qy 539 ggcggatttgaagtcaggtgtgaaactcgggctcaaccctggcctgcacttgaac 598
Db 553 GCGGACATGTAACTCAGGTCTG-AACTCGCGGCTCAACTCGCAGCCTGCACCTGAAAC 611
Qy 599 tacaagtctgagttgggagagcaagtggaattccaggtgtagcggtagaattgcgtag 658
Db 612 TATGTCTCGAGTTTGGGAGAGCAAGTGAATTCAGAGTGTAGCGGTGAATTCGGTAG 671
Qy 659 atactgagaaacaccagtggaagcgacttgcgtgcacaaactgacgtgagga 718
Db 672 ATATCTGAGGAAACACCACTGCGGAAGCGACTTGCCTGCGCTAAACTGACGCTGAGCA 731
Qy 719 cgaaagcctgggttagtaaacgggtagatccccgggtataccccgcctcaaacgtgtc 778
Db 732 CGAAACGCTGGGTACTGAACGGGATAGATACCCCGGTAAATCCACGCCCTCAACGTTGTC 791
Qy 779 taceagttgtgggggttttaacctcagtaacgaacacacgaattagtagccgcct 838
Db 792 TACCAGTTGTTGGGGGTTTTAACCTCTAGTAAAGAA-CTAACGGATTAAGTAGACCGCCT 850
Qy 839 ggggactatgctcgaagtgaaactcaaggaattgacgggggtccgcacaagcggtg 898
Db 851 GCGGACTATGCTCGCAAGAGTGAACCTCAAGGAATTCACGGGGTNNNCACAGCGGTG 910
Qy 899 gggcatgtggttattcgatgatacccccaaaacctcaacctggcgttgacatgagctg 958
Db 911 GAGCATGTGGTTAATTCATGATACGGCGAAACCTCACCTAGGCTTGACATGGAGTGG 970
Qy 959 aatcatgtagatataatgactctcggcagattccacaggtgctgcagtgtgtcgtca 1018
Db 971 AATTATGTAGATATCATGAG-CTTCGGGCGCTTCACAGGTGCTGCATGTTGTCGTCA 1029
Qy 1019 gctcgtgctgtagatgttggttaagtcgccgcaacgcagcgcacccctatogtatgtg 1078
Db 1030 GCTCGTGTCTGAGATGTGGCTTAAGTTCGCCCAACGAGCGCAACCCCTCACCTTATGTTG 1089
Qy 1079 cta-cccttaagttggcactggtacgaactcccggtgacaaaaccggaggaagcgaggga 1137
Db 1090 CCATCATTTAGTTGGGCACTCTGTAAGAACTCCCGGTGACAAACCGGAGGAGCGGGGA 1149
Qy 1138 tgacgtcgaactcctcgtgcttattgtccaggccacacacgctgctacaaatggccgata 1197
Db 1150 TGACGTCAANTCTCATGGCCCTTATGTCTAGGGCAACACACGCTGCTACATGGCCGTA 1209
Qy 1198 cagaggggtcgcaactcgaagaggagctaatctctataaagtcggtccagttcggatt 1257
Db 1210 CAAAGGGTAGCCAACTTCGCGAGGGGAGCTAATCTCAAAAAGCCGCTCCAGTTCCGATT 1269
Qy 1258 ggggtcgaactcgaccccatgaagtgcgaatcactagtaatcgcaatcagcatgcgcg 1317
Db 1270 GGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGGATCAGCATGCCG 1329
Qy 1318 cgggtgaatcgttcccgagccttgtacacacgcgcctgcacacccactgagtgaggagca 1377
Db 1330 CGGTGAATACGTTCCCGGACCTTGTACACACCGCCGCTCACACCCACTCGTGGGGAGCA 1389
```

```
Qy 1378 ccgaaagtgtcttgttaaccgttaaggacagactactactaagtgaacctcgtaaaagg 1437
Db 1390 CCCGAAGTGTCTTTCACCAACCGCAGGAGCAGACTACTAAGTGAACCTCGTGAAGG 1449
Qy 1438 ggtgaagtcgtaacaagggtacc 1459
Db 1450 GGTGAAGTCGTAACAAGGTAGC 1471
```

## RESULT 12

```
LK16SRDNP 1419 bp DNA BCT 02-DEC-1993
LOCUS L.kirschneri gene for 16S ribosomal RNA.
DEFINITION 221628
VERSION 221628.1 GI:433582
KEYWORDS 16S ribosomal RNA.
SOURCE Leptospira kirschneri.
ORGANISM Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE 1 (bases 1 to 1419)
AUTHORS Hooke, J. V.
TITLE Phylogeny of Leptospiraceae and related Spirochaetes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1419)
AUTHORS Hooke, J. V.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-1993) Hooke, J. V., Public Health Laboratory
Service, Leptospira Reference Laboratory, Stonebow Road, Hereford,
Herefordshire, United Kingdom, HR1 2ER
```

## FEATURES

```
source
1..1419
/organism="Leptospira kirschneri"
/strain="CYNOPTERI 3522 C"
/db_xref="taxon:29507"
<1..>1419
```

```
BASE COUNT 369 a 326 c 423 g 299 t 2 others
ORIGIN
```

## Query Match

```
..84.4% Score 1246.6; DB 2; Length 1419;
Best Local Similarity 94.2%; Pred. No. 3.4e-13;
Matches 1336; Conservative 2; Mismatches 76; Indels 5; Gaps 4;
```

```
Qy 16 ctaacgtggcgcgctcttaaacatgc--aagtcagcggggtagcaatcacctagcgg 73
```

```
Db 1 STARCGCTGGCGCGGCTCTTAAACATTCCAAGTCCCAAGCGGAGTAGCAATACTACGCGG 60
```

```
Qy 74 cgaacgggtgagtaacacgt-ggtaactctcctccagctcggataacttccgaaagg 132
```

```
Db 61 CGAACGGTTGAGTAACAGTGGGTAATCTCTCCGAGCTCGGATACATTTCCGAAAGG 120
```

```
Qy 133 aaagctaataccggaagtcctgttggtacacaagatttgataggtaaaagattattgct 192
```

```
Db 121 GAAGCTAATACTGGATGTCGCCGAGAGATCATAGATTTTTCGGGTAAAGATTTATTGCT 180
```

```
Qy 193 tggagatgagcccgccggtgattagctggttgtagtaattgctcaccacgaagcaga 252
```

```
Db 181 CGGAGATGAGCCGCGTCCGATTAGCTAGTTAGTGGTGAAGGCTCAACGAGCGCAGCA 240
```

```
Qy 253 tcggtagcggcctgaagaggtgccggccacaattgggaactgaacacggtccatctcc 312
```

```
Db 241 TCGGTAGCGCGGCTCGAGAGGGTGTTCGGCCACAAATGGAACTGAGACGCGTCCATCTCC 300
```

```
Qy 313 lacggagcgacagcttaagaattctgctcaatgggggaaacccctgaagcgagcgacgcg 372
```

```
Db 301 TACGGAGCGACAGCAGTAAAGAACTTGTCTCAATGGGGGAACCCCTGAAGCAGCGACGCG 360
```

```
Qy 373 cgtgaacgaagaagctcttcggttgtaagtctcattaggca-ggaaaaataagcagcaa 431
```

```
Db 361 COTGAACGATGAAGGCTTTCGGATTGTAAAGTTAGTAAGCGGGAAGGAAATAAGCAGCAA 420
```



Db 561 GCGGACATGTAAGTCAGGTGTG-AAACTGCGGGCTCAACTCGCAGCCTGCACCTTGAAC 619  
Qy 599 tacaagtcgtgagtttgggaagcaagtggaattccaggtgtagcggtagaaatcgtag 658  
Db 620 TATGCTGTGAGTTTGGAGAGCAAGTGAATTCAGGTGTAGCGGTGAATCCGTAG 679  
Qy 659 atactgaggaacacacagtcgcaagcgactgctggtcctcaaaactgacgtgagca 718  
Db 680 ATATCTGGAGGAACACCAAGTGGCAAGGCGACTTGTGGCTTAACACTGACGCTGAGCA 739  
Qy 719 cgaagcgtggttagtaaaacggatttagatacccccggttaaccacgcccctaaacgtgtgc 778  
Db 740 CGAAGCGTGGTAGTAACCGGATTAGATACCCCGGTAAATCCAGCCCTAAACGTTGTC 799  
Qy 779 taccagttgttgggggttttaacctcagtaacgaacactaacgattagtagaccgct 838  
Db 800 TACCAGTTGTTGGGGGTTTTAACCTCAGTAACGAA-CTAACGGATTAAAGTAGACCGCT 858  
Qy 839 gggactatgctcgaagtgaaactcaaaaggaattgacgggggttcgcacaagcggtg 898  
Db 859 GGGGACTATGCTCGACAGAGTGAACCTCAAAAGGAATTGACGGGGGTNNNCACAGCGGTG 918  
Qy 899 gagcatgtggttaattcgtatgatacccaaaaacccctcacctggcttgacatggtatcg 958  
Db 919 GAGCATGTGTTAATTCGATGATACGCGAATAACCTCACCTAGCTTGACATGAGTGG 978  
Qy 959 aatcatgtagagatatagaccttcgggcagatgcacagtgctgcatggttgcgtca 1018  
Db 979 AATCATGTAGAGATACATAGAG-CTTCGGGGCGCTTCACAGGTGCTGATGTTGTCGTCA 1037  
Qy 1019 gctcgtgctgagatgttgggttaagtcgcaacgacgcaacccctatcgtatgtg 1078  
Db 1038 GCTGCTGCTGAGATGTGGGTTAAGTCCCGCAACGAGCGCAACCCCTACCTTATGTTG 1097  
Qy 1079 cta-cccttaagttgggcaactgacgaactgcccgtgacaaaacggaggaagcgggga 1137  
Db 1098 CCATCATTTAGTTGGGCACTCGTAAGGAACCTGCCGTGACAAACCGGAGGAAGCGGGGA 1157  
Qy 1138 tgacgtcaaa tccatgagccttatgttccagggccacacacgctgctacaatggccgata 1197  
Db 1158 TGAGCTNNNNNNNNNNCCCTTTATGTCTATGGCAACACACACGCTGCTACAATGGCCGTA 1217  
Qy 1198 caggggtgcgaactgcgaagaggagcgtactctcaaaagtcggtcccgatcgtcgatt 1257  
Db 1218 CAAGGGTAGCCAACTCCGAGGGGAGCTAATCTCAAAAAGCCGCTCCCAAGTTCGGATT 1277  
Qy 1258 ggggtctgcaactgacccccatgaagtgcgaaatcgctagtaaatcgcgatcagcatgcg 1317  
Db 1278 GGAGTCTCAACTCGACTCCATGAAGTCGGAATCGTAGTAATCGCGGATCAGCATGCGG 1337  
Qy 1318 cgggtgaatacgttcccgacacctgtacacaccccgctcacacacactgagtgggagca 1377  
Db 1338 CGGTGAATACGTTCCCGGACCTTTACACACCCCGCTCACACACCTGAGTGGGGAGCA 1397  
Qy 1378 cccgaagtggcttcttgaacccgtaagagacagactactaagtgaaactcgtaaaagg 1437  
Db 1398 CCCGAGTGGTCTTTGCCAACCCGTAAGGAAGCAGCACTACTAAGGTGAACCTCGTGAAGG 1457  
Qy 1438 ggtgaagtcgtacaaggtacc 1459  
Db 1458 GGTGAAGTCGTAAACAAGGTAGC 1479

## RESULT 14

LM16SRDX  
LOCUS LM16SRDX 1418 bp DNA BCT 02-DEC-1993  
DEFINITION L.meyeri gene for 16S ribosomal RNA (partial).  
ACCESSION Z21648  
VERSION Z21648.1 GI:433585  
KEYWORDS 16S ribosomal RNA.  
SOURCE Leptospira meyeri.  
ORGANISM Bacteria; Spirochaetales; Leptospiraceae; Leptospira.

## REFERENCE

AUTHORS Hookey,J.V.  
TITLE Phylogeny of Leptospiraceae and related Spirochaetes  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1418)  
AUTHORS Hookey,J.V.  
JOURNAL Direct Submission  
TITLE Submitted (09-FEB-1993) Hookey J.V., Public Health Laboratory  
JOURNAL Service, Leptospira Reference Laboratory, Stonebow Road, Hereford,  
Herefordshire, United Kingdom, HR1 2ER

## FEATURES

source

1..1418  
/organism="Leptospira meyeri"  
/strain="RANARUM ICE"  
/db\_xref="taxon:29508"  
<1..>1418  
/product="16S ribosomal RNA"

BASE COUNT 372 a 324 c 421 g 299 t

## ORIGIN

Query Match 83.8%; Score 1238.2; DB 2; Length 1418;  
Best Local Similarity 94.2%; Pred. No. 4.6e-13;  
Matches 1337; Conservative 2; Mismatches 75; Indels 5; Gaps 5;

Qy 16 ctaacgctggcgcgctcttaaacatgcaagtcgagcggggtgacataacctagcgcg 75

Db 1 STARCGTGGCGGCGCTTTAAACATCCAA-TCAAGCGGAGTAGCAATACTACGCGCG 59

Qy 76 aacgggtgagtaaacagt-ggtaatcttccctccagtcctggataaacttccgaagaa 134

Db 60 AACGGGTGAGTAACACGTCGGGTATCTTCCTCCGAGTCTGGGATAACTTTTCGAAGGGA 119

Qy 135 agctaataccggatgtagtctgttgatcacagatgttgatgaagaattattgcttg 194

Db 120 AGCTAATACTGATAGTCCCGAGAGATCATAAAGATTTTTCGGGTAAAGATTTCATGCTTG 179

Qy 195 gagatgagccgcgccgattagctagttggtgaggttaattggctcaccagcgacgac 254

Db 180 GAGATGAGCCCGCTCCGATTAGCTAGTTGGTAGGTATGGCTCACCAAGCGGACGATC 239

Qy 255 ggtagcggctgaggggtgtccggccacaatggaactgagacacgctccatacctcta 314

Db 240 GGTAGCCGCTGAGAGGCTGTTCCGCCACAATGGAACCTGAGACACGCTCCATCTCTA 299

Qy 315 cgggagcgagcagttaagaattcttctcaatgggggaacccctgaagcagcgacg 374

Db 300 CGGAGGCAGCAGTTAAACAATCTTGTCAATGGGGGAACCTTGAAGCAGCGACGCCGCG 359

Qy 375 tgaacgaagaaggtcttcggattgtaaaagtccattagga-ggaaaaataagcagcaatg 433

Db 360 TGAACGATGAAGGCTTTCGGATTGTAAAGTTCAATAAGCAGGGAATAAGCACAATG 419

Qy 434 tgatgattgaccttcctaaagcaccggttaactacgtgcccagcagcgcggttaatacgt 493

Db 420 TGATGATGTTACCTGCCCTAAAGCACCAGCTAACTACGTGCCAGCAGCGCGGTATAGT 479

Qy 494 atggtgcaagcgttcttcggaatcattggcgtaaaaggtgctgtagcggtttgtaagt 553

Db 480 ATGTTGAAGCGTTTTCGGAATCATTTGGCGTAAAGGTTGCTAGCGGACATGTAAGT 539

Qy 554 caggtgtgaaactgcggtcaccggtgcccctgcaactgaaactacaagtctggaatt 613

Db 540 CAGGTGTGAAACCTGGGGGCTCAACTCCAGCCTTCACCTTGAAACTATGTGCTCGAGTT 599

Qy 614 tgggagcgcaagtgggaattccaggtgtagcggtagaaatcgtagatatctcgaggaaca 673

Db 600 TGGGAGAGCAAGTGAATTCAGGTGTAGCGGTGAAATGCGTAGATATCTGGAGACA 659

Qy 674 ccagtgcggaagcgacttctgctcaaaactgacgctgagggcacgaaagcgtgggtag 733

Db 660 CCAGTGGCGAAGGCGACTTGTGTGGCTTAAACTGACGCTAGGCGACGAAAGCGTGGGTAG 719



Db 913 GACATGGAGTGGATCATGTAGATACATGAGCCTTCGGGCCCGCTTCACAGGTGCTGCA 972  
Qy 1007 tggttgtcgtcagctcgtgctcgtgagatgttgggttaagtcccgcaacgagcgcaacccc 1066  
Db 973 TGGTTGTCGTGAGTCGTGTCGTGAGATGTTGGGTAAAGTCCCGCAACGACGCAACCCCT 1032  
Qy 1067 ktcgltatgttgcta-cccttaagtgggactggtgtaagaaactcccggtgacaaacccgga 1125  
Db 1033 CACCTTATGTTGCCATCATTCAGTTGGGCACTCGTAAGGAACTGCCGCTGACAAACCGGA 1092  
Qy 1126 ggaaggggggagatgacgtcaaatcctcctgcttcttattgtccaggggccacacagtgacta 1185  
Db 1093 GGAAGGGGGGATGACGTCAATCCTCATGGCCCTTATGCTAGGGCAACACACGIGCTA 1152  
Qy 1186 caatggccgatacacagaggggtcgccaaactcgcaagaggagcctaactctctaaagtcggtc 1245  
Db 1153 CAATGGCCGGTACAAAGGGTAGCCAACTCGCGAGGGGAGCTAATCTCAAAAAATCCGGTC 1212  
Qy 1246 ccagttcggattgggggtctgcaactcgaccccatgaagtcgggaatcgctagtaatcgcg 1305  
Db 1213 CCAGTTCGGATTGGAGTGTGCAACTCGACTCCATGAAAGTCGGAATCGTATGTAATCGCG 1272  
Qy 1306 atcagcatgccggtgtaatacgttcccgaccttgtacacacgcccgtcacaccacct 1365  
Db 1273 ATCAGCATGCCGCGGTGAAATACGTTCCGGACCTTGTACACACCGCCGCTCACACACCT 1332  
Qy 1366 gagtggggagcaccggaagtgttttttaaccgtaagagacagactactaaggtgaa 1425  
Db 1333 GAGTCGGGAGCACCAGTGGTCTTTGCCAAACCGCAAGCAAGCAGACTACTAAGGTGAA 1392  
Qy 1426 actc 1429  
Db 1393 ACTC 1396

Search completed: March 27, 2001, 08:16:52  
Job time: 5889 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:17:03 ; Search time 2286.42 Seconds  
(without alignments)  
15.668 Million cell updates/sec

Title: US-09-380-826A-4  
Perfect score: 7  
Sequence: 1 tgttga 7

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_om.\*

4: gb\_ov.\*

5: gb\_ph.\*

6: gb\_pl1.\*

7: gb\_pl2.\*

8: gb\_pr1.\*

9: gb\_pr2.\*

10: gb\_pr3.\*

11: gb\_ro.\*

12: gb\_sy.\*

13: gb\_un.\*

14: em\_fun.\*

15: em\_hum1.\*

16: em\_hum2.\*

17: em\_in.\*

18: em\_om.\*

19: em\_or.\*

20: em\_ov.\*

21: em\_pat.\*

22: em\_ph.\*

23: em\_pl.\*

24: em\_ro.\*

25: em\_sts.\*

26: em\_sy.\*

27: em\_un.\*

28: em\_v1.\*

29: gb\_ba3.\*

30: gb\_in1.\*

31: gb\_in2.\*

32: gb\_in3.\*

33: gb\_pl3.\*

34: gb\_pr4.\*

35: em\_ba1.\*

36: em\_ba2.\*

37: em\_htg1.\*

38: em\_htg2.\*

39: em\_htg3.\*

40: em\_htg4.\*

41: em\_htg5.\*

42: em\_htg6.\*

43: em\_htg7.\*

44: em\_htg8.\*  
45: em\_htg9.\*  
46: em\_htg10.\*  
47: em\_hum3.\*  
48: em\_hum4.\*  
49: em\_hum5.\*  
50: em\_hum6.\*  
51: gb\_pr5.\*  
52: gb\_pr6.\*  
53: gb\_pr7.\*  
54: gb\_htg1.\*  
55: gb\_htg2.\*  
56: gb\_htg3.\*  
57: gb\_htg4.\*  
58: gb\_htg5.\*  
59: gb\_htg6.\*  
60: gb\_htg7.\*  
61: gb\_htg8.\*  
62: gb\_htg9.\*  
63: gb\_htg10.\*  
64: gb\_htg11.\*  
65: gb\_htg12.\*  
66: gb\_htg13.\*  
67: gb\_htg14.\*  
68: gb\_htg15.\*  
69: gb\_htg16.\*  
70: gb\_htg17.\*  
71: gb\_htg18.\*  
72: gb\_htg19.\*  
73: gb\_htg20.\*  
74: gb\_htg21.\*  
75: gb\_htg22.\*  
76: gb\_htg23.\*  
77: gb\_sts1.\*  
78: gb\_sts2.\*  
79: gb\_v11.\*  
80: gb\_v12.\*  
81: gb\_pat1.\*  
82: gb\_pat2.\*  
83: em\_htg0.\*  
84: gb\_htg24.\*  
85: gb\_pr8.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	7	100.0	10	81	AR062865	AR062865 Sequence
C 2	7	100.0	13	81	A00432	A00432 Nucleotide
C 3	7	100.0	14	51	AX009010	AX009010 Sequence
C 4	7	100.0	14	79	AX007733	AX007733 Sequence
C 5	7	100.0	14	79	AX007734	AX007734 Sequence
C 6	7	100.0	14	81	A25812	A25812 polynucleot
C 7	7	100.0	14	81	A40518	A40518 Sequence 55
C 8	7	100.0	14	81	A88271	A88271 Sequence 41
C 9	7	100.0	14	81	A88273	A88273 Sequence 42
C 10	7	100.0	14	81	A89045	A89045 Sequence 11
C 11	7	100.0	14	81	A90238	A90238 Sequence 41
C 12	7	100.0	14	81	A90240	A90240 Sequence 42
C 13	7	100.0	14	81	I26234	I26234 Sequence 19
C 14	7	100.0	15	79	AX007732	AX007732 Sequence
C 15	7	100.0	15	81	A88272	A88272 Sequence 42
C 16	7	100.0	15	81	A88466	A88466 Sequence 61
C 17	7	100.0	15	81	A90239	A90239 Sequence 42
C 18	7	100.0	15	81	A90433	A90433 Sequence 61
C 19	7	100.0	15	81	AR033435	AR033435 Sequence
C 20	7	100.0	15	81	AR033480	AR033480 Sequence
C 21	7	100.0	15	81	I57664	I57664 Sequence 20

c 22 7 100.0 15 81 I57709 Sequence 24  
 c 23 7 100.0 16 81 A96937 Sequence 15  
 c 24 7 100.0 16 81 A97828 Sequence 10  
 c 25 7 100.0 16 81 A984439 Sequence  
 c 26 7 100.0 16 81 AR093881 Sequence 32  
 c 27 7 100.0 16 81 I34946 Sequence 15  
 c 28 7 100.0 16 81 I52073 Sequence 43  
 c 29 7 100.0 17 81 A65762 Sequence 18  
 c 30 7 100.0 17 81 A97904 Sequence 18  
 c 31 7 100.0 17 81 AR046017 Sequence  
 c 32 7 100.0 17 81 AR053055 Sequence  
 c 33 7 100.0 17 81 AR05016 Sequence  
 c 34 7 100.0 17 81 I04576 Sequence 6  
 c 35 7 100.0 17 81 I13562 Sequence 4  
 c 36 7 100.0 17 81 I15174 Sequence 23  
 c 37 7 100.0 17 81 I32561 Sequence 25  
 c 38 7 100.0 17 81 I37420 Sequence 43  
 c 39 7 100.0 17 81 I50672 Sequence 4  
 c 40 7 100.0 17 81 I53069 Sequence 81  
 c 41 7 100.0 17 82 I89345 Sequence 2  
 c 42 7 100.0 17 82 I94270 Sequence 43  
 c 43 7 100.0 18 81 A07301 Nucleotide  
 c 44 7 100.0 18 81 A07302 Nucleotide  
 c 45 7 100.0 18 81 A45212 Sequence 89

## ALIGNMENTS

RESULT 1  
 LOCUS AR062865/c 10 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 13 from patent US 5843767.  
 ACCESSION AR062865  
 VERSION AR062865.1 GI:5990556  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 10)  
 AUTHORS Beattie, K. L.  
 TITLE Microfabricated, flowthrough porous apparatus for discrete detection of binding reactions  
 JOURNAL Patent: US 5843767-A 13 01-DEC-1998;  
 FEATURES Location/Qualifiers  
 1..10  
 /organism="unknown"  
 BASE COUNT 4 a 4 c 0 g 2 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 t g t t g g a 7  
 Db 7 T G T T G G A 1

RESULT 2  
 LOCUS A00432 13 bp DNA PAT 11-FEB-1993  
 DEFINITION Nucleotide sequence 7 from patent number WO9010459.  
 ACCESSION A00432  
 VERSION A00432.1 GI:14504  
 KEYWORDS unidentified.  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 13)  
 AUTHORS  
 TITLE PHARMACEUTICAL COMPOSITION USEFUL IN THE PREVENTION OR TREATMENT OF

JOURNAL PAPILLOMAVIRUS-INDUCED TUMOURS  
 Patent: WO 9010459-A 7 20-SEP-1990;  
 Transgene S.A  
 FEATURES Location/Qualifiers  
 1..13  
 /organism="unidentified"  
 /db\_xref="taxon:32844" 2 t  
 BASE COUNT 4 a 3 c 4 g  
 ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 t g t t g g a 7  
 Db 10 T G T T G G A 4

RESULT 3  
 LOCUS AX009010 14 bp DNA PRI 06-SEP-2000  
 DEFINITION Sequence 43 from Patent WO9963975.  
 ACCESSION AX009010  
 VERSION AX009010.1 GI:9996384  
 KEYWORDS human.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Brysch, W., Schlingensiepen, K. H. and Schlingensiepen, R.  
 TITLE A method for stimulating the immune system  
 JOURNAL Patent: WO 9963975-A 16-DEC-1999;  
 BIOGOSTIK GES (DE); BRYSCH WOLFGANG (DE); SCHLINGENSIEPEN KARL  
 HERMANN (DE); SCHLINGENSIEPEN REIMAR (DE)

FEATURES Location/Qualifiers  
 1..14  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 3 a 3 c 4 g 4 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 51; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 5e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 t g t t g g a 7  
 Db 3 T G T T G G A 9

RESULT 4  
 LOCUS AX007733/c 14 bp DNA VRL 06-SEP-2000  
 DEFINITION Sequence 275 from Patent WO9967428.  
 ACCESSION AX007733  
 VERSION AX007733.1 GI:9995430  
 KEYWORDS Aids-associated retrovirus.  
 SOURCE Aids-associated retrovirus.

ORGANISM Aids-associated retrovirus  
 Viruses; Retroold viruses; Retroviridae.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Stuyver, L.  
 TITLE Method for detection of drug-selected mutations in the hiv protease gene

JOURNAL Patent: WO 9967428-A 29-DEC-1999;  
 INNOGENETICS NV (BE); STUYVER LIEVEN (BE)  
 FEATURES Location/Qualifiers  
 1..14  
 /organism="Aids-associated retrovirus"



```

BASE COUNT      5 a      /db_xref="taxon:11966"      3 t
ORIGIN          5 a      6 c      0 g

Query Match      100.0%; Score 7; DB 79; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 12 TGTGGA 6

RESULT 5
LOCUS AX007734/c      14 bp      DNA      06-SEP-2000
DEFINITION Sequence 276 from Patent WO9967428.
ACCESSION AX007734
VERSION AX007734.1 GI:9995431
KEYWORDS Aids-associated retrovirus.
SOURCE Aids-associated retrovirus.
ORGANISM Aids-associated retrovirus
REFERENCE 1 (bases 1 to 14)
AUTHORS Stuyver, L.
TITLE Method for detection of drug-selected mutations in the hiv protease
gene.
JOURNAL Patent: WO 9967428-A 29-DEC-1999;
INNOGENETICS NV (BE); STUYVER LIEVEN (BE)
FEATURES Location/Qualifiers
source 1..14
/organism="Aids-associated retrovirus"
/db_xref="taxon:11966"
BASE COUNT      5 a      6 c      0 g      3 t
ORIGIN

Query Match      100.0%; Score 7; DB 79; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 13 TGTGGA 7

RESULT 6
LOCUS A25812/c      14 bp      DNA      14-MAR-1995
DEFINITION polynucleotide 14C12.
ACCESSION A25812
VERSION A25812.1 GI:904780
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 14)
AUTHORS
JOURNAL Patent: FR 2680520-A 19 26-FEB-1993;
FEATURES Location/Qualifiers
source 1..14
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT      4 a      4 c      2 g      4 t
ORIGIN

Query Match      100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7

```

```

Db 12 TGTGGA 6

RESULT 7
LOCUS A40518      14 bp      DNA      05-MAR-1997
DEFINITION Sequence 55 from Patent WO9425578.
ACCESSION A40518
VERSION A40518.1 GI:2296553
KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 14)
AUTHORS
TITLE ANTISENSE-OLIGONUCLEOTIDES FOR THE TREATMENT OF IMMUNOSUPPRESSIVE
EFFECTS OF TRANSFORMING GROWTH FACTOR--g(b) (TGF--g(b))
JOURNAL Patent: WO 9425578-A 55 10-NOV-1994;
BIOGNOSTIK GES (DE)
FEATURES Location/Qualifiers
source 1..14
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      3 a      3 c      4 g      4 t
ORIGIN

Query Match      100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 3 TGTGGA 9

RESULT 8
LOCUS A88271/c      14 bp      DNA      22-JAN-2000
DEFINITION Sequence 419 from Patent WO9833904.
ACCESSION A88271
VERSION A88271.1 GI:6736841
KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Brysch, W. and Schlingensiepen, K.
TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD
JOURNAL Patent: WO 9833904-A 06-AUG-1998;
BIOGNOSTIK GES (DE); BRYSCH WOLFGANG (DE)
FEATURES Location/Qualifiers
source 1..14
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      5 a      4 c      0 g      5 t
ORIGIN

Query Match      100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7
Db 11 TGTGGA 5

RESULT 9
LOCUS A88273/c      14 bp      DNA      22-JAN-2000
DEFINITION Sequence 421 from Patent WO9833904.

```

ACCESSION A88273  
VERSION A88273.1 GI:6736843  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 14)  
AUTHORS Brysch, W. and Schlöngensiepen, K.  
TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD  
JOURNAL Patent: WO 9833904-A 06-AUG-1998;  
BIOGOSTIK GES (DE); BRYSCH WOLFGANG (DE)  
FEATURES Location/Qualifiers  
source 1..14  
/organism="unidentified"  
/db\_xref="taxon:32644" 4 t  
BASE COUNT 6 a 4 c 0 g  
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttga 7  
|||||  
Db 12 TGTGGA 6

RESULT 10  
LOCUS A89045 14 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 1193 from Patent WO9833904.  
ACCESSION A89045  
VERSION A89045.1 GI:6737615  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 14)  
AUTHORS Brysch, W. and Schlöngensiepen, K.  
TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD  
JOURNAL Patent: WO 9833904-A 06-AUG-1998;  
BIOGOSTIK GES (DE); BRYSCH WOLFGANG (DE)  
FEATURES Location/Qualifiers  
source 1..14  
/organism="unidentified"  
/db\_xref="taxon:32644" 4 t  
BASE COUNT 3 a 3 c 4 g  
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttga 7  
|||||  
Db 3 TGTGGA 9

RESULT 11  
LOCUS A90238 14 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 419 from Patent EP0856579.  
ACCESSION A90238  
VERSION A90238.1 GI:6738752  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 14)  
AUTHORS Brysch, W. D. and Schlöngensiepen, K. D.  
TITLE An antisense oligonucleotide preparation method

JOURNAL Patent: EP 0856579-A 05-AUG-1998;  
BIOGOSTIK GES (DE)  
FEATURES Location/Qualifiers  
source 1..14  
/organism="unidentified"  
/db\_xref="taxon:32644" 5 t  
BASE COUNT 5 a 4 c 0 g  
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttga 7  
|||||  
Db 11 TGTGGA 5

RESULT 12  
LOCUS A90240/c 14 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 421 from Patent EP0856579.  
ACCESSION A90240  
VERSION A90240.1 GI:6738754  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 14)  
AUTHORS Brysch, W. D. and Schlöngensiepen, K. D.  
TITLE An antisense oligonucleotide preparation method  
JOURNAL Patent: EP 0856579-A 05-AUG-1998;  
BIOGOSTIK GES (DE)  
FEATURES Location/Qualifiers  
source 1..14  
/organism="unidentified"  
/db\_xref="taxon:32644" 4 t  
BASE COUNT 6 a 4 c 0 g  
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttga 7  
|||||  
Db 12 TGTGGA 6

RESULT 13  
LOCUS I26234/c 14 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 19 from patent US 5556955.  
ACCESSION I26234  
VERSION I26234.1 GI:1606104  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 14)  
AUTHORS Verghnaud, G.  
TITLE Process for detection of new polymorphic loci in a DNA sequence,  
nucleotide sequences forming hybridization probes and their  
applications  
JOURNAL Patent: US 5556955-A 19 17-SEP-1996;  
FEATURES Location/Qualifiers  
source 1..14  
/organism="unknown" 4 t  
BASE COUNT 4 a 4 c 2 g  
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgga 7  
| | | | |  
Db 12 TGTGGA 6

RESULT 14  
AX007732/c  
LOCUS AX007732 15 bp DNA VRL 06-SEP-2000  
DEFINITION Sequence 274 from Patent WO967428.  
ACCESSION AX007732  
VERSION AX007732.1 GI:9995429  
KEYWORDS  
SOURCE Aids-associated retrovirus.  
ORGANISM Aids-associated retrovirus.  
REFERENCE 1 (bases 1 to 15)  
AUTHORS Stuyver, L.  
TITLE Method for detection of drug-selected mutations in the hiv protease gene  
JOURNAL Patent: WO 967428-A 29-DEC-1999;  
INNOGENETICS NV (BE); STUYVER LIEVEN (BE)  
FEATURES  
source  
1..15  
Location/Qualifiers  
/organism="Aids-associated retrovirus"  
/db\_xref="taxon:11966"  
BASE COUNT 6 a 6 c 0 g 3 t  
ORIGIN

Query Match 100.0%; Score 7; DB 79; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgga 7  
| | | | |  
Db 13 TGTGGA 7

RESULT 15  
A88272/c  
LOCUS A88272 15 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 420 from Patent WO9833904.  
ACCESSION A88272  
VERSION A88272.1 GI:6736842  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 15)  
AUTHORS Brysch, W. and Schlingensiepen, K.  
TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD  
JOURNAL Patent: WO 9833904-A 06-AUG-1998;  
BIOGNOSTIK GES (DE); BRYSCH WOLFGANG (DE)  
FEATURES  
source  
1..15  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 6 a 4 c 0 g 5 t  
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgga 7  
| | | | |  
Db 12 TGTGGA 6



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:22:58 ; Search time 207.51 Seconds  
(without alignments)  
12.672 Million cell updates/sec

Title: US-09-380-826A-4

Perfect score: 7

Sequence: 1 tgttga 7

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 18781343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

```
1: /cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	7	100.0	7	19	V58898 Leptospira nucleot
2	7	100.0	10	20	X99938 Human parkin gene
3	7	100.0	10	21	Z78251 Human dendritic ce
4	7	100.0	10	21	Z80950 Metastatic breast
5	7	100.0	10	21	Z82660 Metastatic breast
6	7	100.0	10	21	Z84297 Metastatic breast
7	7	100.0	11	20	X77658 NII active EGS 22.
8	7	100.0	14	15	Q40606 Hypervariable regi
9	7	100.0	14	15	Q78406 Antisense oligonuc
10	7	100.0	14	19	V97198 Potato citrate syn
11	7	100.0	14	19	V48830 ErbB-2 gene antise
12	7	100.0	14	19	V48832 ErbB-2 gene antise

13	7	100.0	14	20	A21661 Integrin alpha 6 s
14	7	100.0	14	20	V92044 Human C-raf target
15	7	100.0	14	21	Z97785 HIV-1 protease gen
16	7	100.0	14	21	Z97786 HIV-1 protease gen
17	7	100.0	14	21	Z65483 Immunosuppressant
18	7	100.0	14	21	Z64788 Substrate for hair
19	7	100.0	15	2	N10044 Sequence of initia
20	7	100.0	15	2	N10045 Sequence of initia
21	7	100.0	15	17	X66645 Human CD40 hammerh
22	7	100.0	15	17	X66275 Mouse B7-2 hammerh
23	7	100.0	15	17	X66276 Mouse B7-2 hammerh
24	7	100.0	15	17	X66274 Mouse B7-2 hammerh
25	7	100.0	15	17	X66235 Mouse B7-2 hammerh
26	7	100.0	15	17	X66236 Mouse B7-2 hammerh
27	7	100.0	15	17	X66237 Mouse B7-2 hammerh
28	7	100.0	15	17	X66238 Mouse B7-2 hammerh
29	7	100.0	15	17	X65872 Human B7-2 hammerh
30	7	100.0	15	17	X65873 Human B7-2 hammerh
31	7	100.0	15	17	Z50248 Rabbit CERP HH rib
32	7	100.0	15	17	Z50250 Rabbit CERP HH rib
33	7	100.0	15	19	V49025 rb gene antisense
34	7	100.0	15	19	V48831 ErbB-2 gene antise
35	7	100.0	15	20	X31673 Tag sequence of a
36	7	100.0	15	21	Z59285 Human NR8 gene pro
37	7	100.0	15	21	Z59298 Human NR8 gene pro
38	7	100.0	15	21	Z59299 Human NR8 gene pro
39	7	100.0	15	21	Z90823 Human NR8 gene pro
40	7	100.0	15	21	Z90839 Human NR8 gene pro
41	7	100.0	15	21	Z90896 Human NR8 gene pro
42	7	100.0	15	21	Z90905 Human NR8 gene pro
43	7	100.0	15	21	Z97784 HIV-1 protease gen
44	7	100.0	15	21	Z62590 Substrate for HH r
45	7	100.0	15	21	Z62635 Substrate for HH r

#### ALIGNMENTS

RESULT 1

V58898

ID V58898 standard; DNA; 7 BP.

XX

AC V58898;

XX

DT 20-JAN-1999 (first entry)

XX

DE Leptospira nucleotide sequence.

XX

KW Infection; pathogenic Leptospira; protective immunity; therapy;

KW diagnosis; ss.

XX

OS Leptospira sp.

XX

PN WO9840099-A1.

XX

PD 17-SEP-1998.

XX

PF 06-MAR-1998; 98WO-AU00145.

XX

PR 07-MAR-1997; 97AU-0005494.

XX

PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

PA (PIGR-) PIG RES & DEV CORP.

XX

PI Chappel RJ;

XX

DR WPI; 1998-520791/44.

XX

PT New isolated pathogenic Leptospira bacterium - useful for, e.g

PT developing products for conferring protective immunity, and for

PT prophylactic or therapeutic treatment

XX

PS Claim 15; Page 71; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar Hurstbridge or the species  
 CC L. fainei. The LS bacteria can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 SQ Sequence 7 BP; 1 A; 0 C; 3 G; 3 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.2e+07;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7  
 Db 1 tgttggg 7

RESULT 2  
 X99938/C  
 ID X99938 standard; DNA; 10 BP.

XX AC X99938;  
 XX 21-OCT-1999 (first entry)  
 DT Human parkin gene intron 7 fragment.

DE Parkinson's disease related gene; parkin gene; variant; gene therapy;  
 KW intron; ss.

XX OS Homo sapiens.

XX KW WO9940191-A1.

XX PN 12-AUG-1999.

XX PD 09-FEB-1999; 99WO-JP00545.

XX PF 09-FEB-1998; 98JP-0027531.

XX PR (MIZU)/ MIZUNO Y.

XX PA (SHIM)/ SHIMIZU N.

XX PI Mizuno Y, Shimizu N;

XX PT WPI; 1999-494295/41.

XX PS Gene implicated in the pathology of Parkinson's disease, used for

XX PT treatment of the disease

XX PS Claim 11; Page 100; 114pp; English.

XX CC This sequence represents a fragment of an intron from the gene of the

XX CC invention. The gene has been designated the parkin gene, and variants of

XX CC it are implicated in the pathology of Parkinson's disease, and found in

XX CC Parkinson's disease patients. The sequences may be used for the

XX CC diagnosis, treatment (including gene therapy) and investigation of

XX CC Parkinson's disease.

XX SQ Sequence 10 BP; 3 A; 3 C; 1 G; 3 T; 0 other;

Query Match 100.0%; Score 7; DB 20; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7  
 Db 1 tgttggg 7

Db 9 TGTGGA 3

RESULT 3

ID Z78251 standard; DNA; 10 BP.

XX AC Z78251;

XX DT 10-APR-2000 (first entry)

XX DE Human dendritic cell SAGE tag, SEQ ID NO:679.

XX KW SAGE tag: serial analysis of gene expression; antigen-presenting cell;

XX KW APC; monocyte-derived dendritic cell; differential gene expression;

XX KW immunostimulatory cofactor; costimulatory factor; CTL;

XX KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.

XX OS Homo sapiens.

XX PN WO9965924-A2.

XX PD 23-DEC-1999.

XX PF 18-JUN-1999; 99WO-US13800.

XX PR 19-JUN-1998; 98US-0089833.

XX PR 19-JUN-1998; 98US-0089844.

XX PR 19-JUN-1998; 98US-0089853.

XX PR 19-JUN-1998; 98US-0089878.

XX PR 19-JUN-1998; 98US-0089991.

XX PR 19-JUN-1998; 98US-0089992.

XX PR 19-JUN-1998; 98US-0089993.

XX PR 19-JUN-1998; 98US-0089994.

XX PR 19-JUN-1998; 98US-0089997.

XX PR 19-JUN-1998; 98US-0089999.

XX PR 19-JUN-1998; 98US-0090000.

XX PR 19-JUN-1998; 98US-0090035.

XX PR 19-JUN-1998; 98US-0090036.

XX PR 19-JUN-1998; 98US-0090039.

XX PR 19-JUN-1998; 98US-0090040.

XX PR 19-JUN-1998; 98US-0090041.

XX PR 19-JUN-1998; 98US-0090042.

XX PR 19-JUN-1998; 98US-0090043.

XX PR 19-JUN-1998; 98US-0090044.

XX PR 19-JUN-1998; 98US-0090045.

XX PR 19-JUN-1998; 98US-0090047.

XX PR 19-JUN-1998; 98US-0090048.

XX PR 19-JUN-1998; 98US-0090072.

XX PR 19-JUN-1998; 98US-0090076.

XX PR 19-JUN-1998; 98US-0090077.

XX PR 19-JUN-1998; 98US-0090078.

XX PR 19-JUN-1998; 98US-0090079.

XX PR 19-JUN-1998; 98US-0090080.

XX PR 08-DEC-1998; 98US-0111715.

XX PA (GENZ ) GENZYME CORP.

XX PA (ROBE/) ROBERTS B L.

XX PA (SHAN/) SHANKARA S.

XX PI Roberts BL, Shankara S;

XX DR WPI; 2000-106077/09.

XX PT Isolated polynucleotides differentially expressed in antigen-presenting

XX PT cells, useful in gene vaccines against cancer -

XX PS Claim 1; Page 84; 130pp; English.

XX CC Sequences Z77573-Z79709 represent SAGE (serial analysis of gene

XX CC expression) tags used to identify mRNA transcripts encoding

XX CC immunostimulatory cofactor proteins which are preferentially or

XX CC differentially expressed in monocyte-derived dendritic cells compared

CC with monocytes. Some of the transcripts correspond to known genes or  
 CC ESTs (expressed sequence tags) which were previously unknown to be  
 CC preferentially or differentially expressed in dendritic cells, while  
 CC other transcripts correspond to novel genes. Antigen-presenting cell  
 CC (APC)-associated costimulatory factors play an important role in the  
 CC activation of the cytotoxic immune response, particularly against tumour  
 CC cells. Tumour antigen presentation via the MHC (major histocompatibility  
 CC complex) and subsequent recognition by T-cell receptors is alone  
 CC insufficient to activate a robust cytotoxic immune response that can lyse  
 CC the tumour cells, immunostimulatory cofactors also being required for  
 CC efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid  
 CC sequences identified using the SAGE tags have several potential uses.  
 CC They may be used in vaccines to induce an immune response, particularly  
 CC against a tumour antigen; to modulate the genotype of an APC; to screen  
 CC for agents that modulate expression of differentially expressed genes in  
 CC an APC; and as hybridisation probes/amplification primers for the  
 CC diagnosis, prognosis and monitoring of diseases related to abnormal  
 CC expression of these genes. Detection of the dendritic cell differentially  
 CC expressed genes, or of their encoded proteins, can be used to identify  
 CC cells as belonging to the monocyte lineage. Cells containing these genes  
 CC can be used in active immunotherapy (or to stimulate production of a  
 CC population of antigen-specific effector cells) and vectors containing  
 CC them are used in gene therapy. Co-administration of tumour antigens and  
 CC APC-associated costimulatory factors ensures adequate antigen  
 CC presentation to endogenous APCs and upregulates the APCs for the  
 CC presentation of co-stimulatory signals, migration to T cell-rich sites,  
 CC secretion of T cell growth factors and secretion of chemokines for  
 CC recruitment of immune effector cells.  
 CC  
 CC Sequence 10 BP; 1 A; 2 C; 3 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
 Db 3 tgttggg 9

RESULT 4  
 ID 280950 standard; DNA; 10 BP.  
 AC 280950;

07-APR-2000 (first entry)

DE Metastatic breast tumour cell upregulated transcript tag #184.

KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
 KW antimetastatic; vaccine; diagnosis; ss.

OS Homo sapiens.

XX WO9965928-A2.

PN 23-DEC-1999.

XX 18-JUN-1999; 99WO-US13647.

XX 19-JUN-1998; 98US-0089853.

PR 19-JUN-1998; 98US-0089997.

PR 19-JUN-1998; 98US-0090039.

PR 19-JUN-1998; 98US-0090040.

PR 19-JUN-1998; 98US-0090041.

XX (GENZ ) GENZYME CORP.

PA (ROBE/) ROBERTS B L.

PA (SHAN/) SHANKARA S.

XX

PI Roberts BL, Shankara S;

XX WPI; 2000-106079/09.

XX Isolated polynucleotides differentially expressed between metastatic  
 PT and non-metastatic breast cancer cells, useful for diagnosis,  
 PT prevention and treatment of cancer -

XX Claim 1; Page 63; 219pp; English.

XX 280767 to 283941 represent tags corresponding to distinct transcripts  
 CC that are preferentially transcribed in the metastatic breast tumour  
 CC tissue (i.e. are upregulated in metastatic breast tumour cells). 283942  
 CC to 286677 represent tags corresponding to distinct transcripts that are  
 CC preferentially transcribed in the primary or non-metastatic breast tumour  
 CC tissue (i.e. are downregulated in metastatic breast tumour cells).  
 CC These transcripts can be used for diagnosis, prognosis, monitoring and  
 CC treatment of breast cancer, particularly where metastatic. Diagnosis is  
 CC by standard immunoassays or hybridisation/amplification reactions.  
 CC Compounds that modulate expression of the transcripts are potentially  
 CC useful for treatment of (metastatic) breast cancer, while promoters from  
 CC the transcripts are used to direct expression, in selected cell types,  
 CC of e.g. therapeutic genes (also ribozymes or antisense sequences),  
 CC particularly an antigen-encoding sequence for use in gene or cell-based  
 CC vaccines. Polypeptides encoded by the transcripts are also useful in  
 CC vaccines; for diagnosing breast cancer and for raising specific  
 CC antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic  
 CC agents. Host cells that produce the polypeptides can be used to expand  
 CC and isolate populations of educated, antigen-specific immune effector  
 CC cells, e.g. cytotoxic T lymphocytes, and these used for adoptive  
 CC immunotherapy.

XX Sequence 10 BP; 2 A; 0 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
 Db 4 tgttggg 10

RESULT 5  
 ID 282660 standard; DNA; 10 BP.  
 XX 282660;

AC 282660;

XX 07-APR-2000 (first entry)

DE Metastatic breast tumour cell upregulated transcript tag #1894.

XX Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
 KW antimetastatic; vaccine; diagnosis; ss.

OS Homo sapiens.

XX WO9965928-A2.

PN 23-DEC-1999.

XX 18-JUN-1999; 99WO-US13647.

XX 19-JUN-1998; 98US-0089853.

PR 19-JUN-1998; 98US-0089997.

PR 19-JUN-1998; 98US-0090039.

PR 19-JUN-1998; 98US-0090040.

PR 19-JUN-1998; 98US-0090041.

XX (GENZ ) GENZYME CORP.

```

PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
PI Roberts BL, Shankara S;
XX WPI: 2000-106079/09.
XX
XX Isolated polynucleotides differentially expressed between metastatic
XX and non-metastatic breast cancer cells, useful for diagnosis,
XX prevention and treatment of cancer -
XX Claim 1: Page 110; 219pp; English.
XX
XX 280767 to 283941 represent tags corresponding to distinct transcripts
XX that are preferentially transcribed in the metastatic breast tumour
XX tissue (i.e. are upregulated in metastatic breast tumour cells). 283942
XX to 286677 represent tags corresponding to distinct transcripts that are
XX preferentially transcribed in the primary or non-metastatic breast tumour
XX tissue (i.e. are downregulated in metastatic breast tumour cells).
XX These transcripts can be used for diagnosis, prognosis, monitoring and
XX treatment of breast cancer, particularly where metastatic. Diagnosis is
XX by standard immunoassays or hybridisation/amplification reactions.
XX Compounds that modulate expression of the transcripts are potentially
XX useful for treatment of (metastatic) breast cancer, while promoters from
XX the transcripts are used to direct expression, in selected cell types,
XX of e.g. therapeutic genes (also ribozymes or antisense sequences),
XX particularly an antigen-encoding sequence for use in gene or cell-based
XX vaccines. Polypeptides encoded by the transcripts are also useful in
XX vaccines; for diagnosing breast cancer and for raising specific
XX antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
XX agents. Host cells that produce the polypeptides can be used to expand
XX and isolate populations of educated, antigen-specific immune effector
XX cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
XX immunotherapy.
XX
XX Sequence 10 BP; 2 A; 0 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgttggga 7
Db 2 tgttggga 8

RESULT 6
284297 ID 284297 standard; DNA; 10 BP.
XX AC 284297;
XX DT 07-APR-2000 (first entry)
XX DE Metastatic breast tumour cell downregulated transcript tag #3531.
XX
XX Human; metastatic breast tumour tissue; breast cancer; tag; primer;
XX non-metastatic breast tumour tissue; gene therapy; anticancer;
XX antimetastatic; vaccine; diagnosis; ss.
XX
XX Homo sapiens.
XX
XX WO965928-A2.
XX
XX 23-DEC-1999.
XX
XX 18-JUN-1999; 99WO-US13647.
XX
XX 19-JUN-1998; 98US-0083853.
XX 19-JUN-1998; 98US-0089997.
XX 19-JUN-1998; 98US-0090039.
XX 19-JUN-1998; 98US-0090040.
XX
PR
```

---

```

PR 19-JUN-1998; 98US-0090041.
XX (GENZ ) GENZYME CORP.
XX (ROBE/) ROBERTS B L.
XX (SHAN/) SHANKARA S.
XX
XX Roberts BL, Shankara S;
XX WPI: 2000-106079/09.
XX
XX Isolated polynucleotides differentially expressed between metastatic
XX and non-metastatic breast cancer cells, useful for diagnosis,
XX prevention and treatment of cancer -
XX Claim 1: Page 153; 219pp; English.
XX
XX 280767 to 283941 represent tags corresponding to distinct transcripts
XX that are preferentially transcribed in the metastatic breast tumour
XX tissue (i.e. are upregulated in metastatic breast tumour cells). 283942
XX to 286677 represent tags corresponding to distinct transcripts that are
XX preferentially transcribed in the primary or non-metastatic breast tumour
XX tissue (i.e. are downregulated in metastatic breast tumour cells).
XX These transcripts can be used for diagnosis, prognosis, monitoring and
XX treatment of breast cancer, particularly where metastatic. Diagnosis is
XX by standard immunoassays or hybridisation/amplification reactions.
XX Compounds that modulate expression of the transcripts are potentially
XX useful for treatment of (metastatic) breast cancer, while promoters from
XX the transcripts are used to direct expression, in selected cell types,
XX of e.g. therapeutic genes (also ribozymes or antisense sequences),
XX particularly an antigen-encoding sequence for use in gene or cell-based
XX vaccines. Polypeptides encoded by the transcripts are also useful in
XX vaccines; for diagnosing breast cancer and for raising specific
XX antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic
XX agents. Host cells that produce the polypeptides can be used to expand
XX and isolate populations of educated, antigen-specific immune effector
XX cells, e.g. cytotoxic T lymphocytes, and these used for adoptive
XX immunotherapy.
XX
XX Sequence 10 BP; 2 A; 0 C; 3 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgttggga 7
Db 2 tgttggga 8

RESULT 7
X77658/c ID X77658 standard; DNA; 11 BP.
XX AC X77658;
XX DT 09-AUG-1999 (first entry)
XX DE N11 active EGS 22.
XX
XX External guide sequence; EGS; target mRNA; identification; diagnostic;
XX inactivation; essential gene; therapy; ss.
XX
XX Synthetic.
XX
XX WO9927135-A2.
XX
XX 03-JUN-1999.
XX
XX 20-NOV-1998; 98WO-US24854.
XX
XX 30-MAR-1998; 98US-0079851.
XX 21-NOV-1997; 97US-0976220.
XX
PR
```



XX (INNO-) INNOVIR LAB INC.  
 XX Kindt TJ, Nilsen TW, Robertson HD;  
 XX WPI; 1999-357853/30.  
 XX Identifying and inhibiting functional nucleic acid molecules in  
 XX cells  
 XX Example 3; Page 28; 58pp; English.  
 XX This invention describes a novel method allowing essential or functional  
 XX genes to be rapidly identified and inactivated. The method is able to  
 XX firstly identify most of the essential genes in an organism (i.e. a  
 XX bacteria or a eukaryote) needed for survival, and secondly it provides  
 XX for reducing or inactivating their expression. The method is able to  
 XX identify functional oligonucleotide molecules able to be used as  
 XX diagnostic reagents and therapeutics. The method provides a means for  
 XX identifying essential genes whose sequence is known only as part of a  
 XX genome with unknown function, as well as a means for identifying  
 XX functional oligonucleotide molecules. The method involves the use of a  
 XX nucleic acid molecule comprising (a) a first reporter gene encoding a  
 XX fusion protein comprising a protein of interest (itself translated from  
 XX an RNA of interest) and a reporter protein, a second reporter gene  
 XX encoding a second reporter protein, and (c) a targeting gene encoding a  
 XX functional oligonucleotide molecule such as an external guide sequence  
 XX (EGS), a ribozyme or an antisense RNA and targeted to the RNA of interest  
 XX at a site on the first reporter gene able to encode the RNA of interest.  
 XX Sequence 11 BP; 4 A; 4 C; 1 G; 2 T; 0 other;

Query Match 100.0%; Score 7; DB 20; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
 |||||  
 Db 10 TGTGGGA 4

RESULT 8  
 Q40606/C  
 ID Q40606 standard; DNA; 14 BP.

XX Q40606;  
 XX 10-AUG-1993 (first entry)

DE Hypervariable region detection probe 14C12.

XX HVR; human; animal; forensic science; paternity testing; diagnosis;  
 XX animal breeding; hereditary diseases; tumours; allele; loss;  
 KW chromosomal regions; tumour region identification; ss.

XX Synthetic.

XX FR2680520-A.

XX 26-FEB-1993.

XX 22-AUG-1991; 91FR-0010516.

XX 22-AUG-1991; 91FR-0010516.

XX (ETFR ) ETAT FRANCAIS.

XX Vergnaud G;

XX WPI; 1993-136548/17.

XX Detecting the hypervariable regions of DNA for diagnosing

PT hereditary illnesses and tumours - by hybridising labelled  
 PT polynucleotides and analysing genomic DNA of individuals which  
 XX react with restriction fragments  
 XX Example; Page 13; 46pp; French.  
 XX The sequence is that of a polynucleotide probe which may be used in  
 CC the detection of new hypervariable regions (HVR) in a DNA sequence.  
 CC HVR represent a fingerprint useful in e.g. forensic science,  
 CC paternity testing, animal breeding, etc. The probe may be used as  
 CC part of a method for the efficient detection in humans or other  
 CC animals, without the use of mini-satellites or primary enrichment.  
 XX Sequence 14 BP; 4 A; 4 C; 2 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 14; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
 |||||  
 Db 12 TGTGGGA 6

RESULT 9  
 Q78406  
 ID Q78406 standard; DNA; 14 BP.

XX Q78406;

XX 27-JUN-1995 (first entry)

DE Antisense oligonucleotide hybridising to TGF-beta gene.

XX Transforming growth factor beta; TGF-beta; antisense; treatment;  
 KW tumour; angiogenesis; breast tumour; neurofibroma; glioma;  
 KW glioblastoma; carcinogenesis; carcinoma; oesophagus; oesophageal;  
 KW gastric; gut; immunosuppression; oligonucleotide; ss.

XX Synthetic.

XX WO9425588-A.

XX 10-NOV-1994.

XX 29-APR-1994; 94WO-EP01362.

XX 30-APR-1993; 93EP-0107089.

XX 13-MAY-1993; 93EP-0107849.

(BIOC-) BIOGNOSTIK GES BIOMOLEKULARE DIAGNOSTIK.

XX Bogdahn U, Brysch W, Schlingensiepen G, Schlingensiepen K;  
 XX Schlingensiepen R;

XX WPI; 1994-358266/44.

XX New transforming growth factor beta anti-sense  
 PT oligo:nucleotide(s) - for treating immunosuppression, tumours,  
 PT etc.

XX Claim 6; Page 40; 74pp; English.

XX The antisense oligonucleotides are useful in the treatment of  
 CC tumours in which expression of TGF-beta is of relevance for  
 CC pathogenicity and/or inhibition of pathological angiogenesis. They  
 CC are used especially for the treatment of the immunosuppressive  
 CC effect of TGF-beta, augmentation of the proliferation of cytotoxic  
 CC lymphocytes, treatment of endogenous hyperexpression of TGF-beta,  
 CC treatment of breast tumours, neurofibromas and malignant gliomas,  
 CC including glioblastomas, treatment and prophylaxis of skin  
 CC carcinogenesis, and treatment of oesophageal and gastric carcinomas.

Wed Mar 28 14:03:18 2001

CC ribozymes can be targeted to specific genes or to consensus sequences  
 CC within a family of related genes, and being catalytic need to be  
 CC present at only very low concentrations.  
 XX Sequence 14 BP; 4 A; 4 C; 5 G; 1 U; 0 other;  
 SQ Sequence 14 BP; 4 A; 4 C; 5 G; 1 U; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
 DB 13 TGTGGA 7

RESULT 11  
 V48830/c  
 ID V48830 standard; RNA: 14 BP.  
 XX V48830;  
 AC V48830;  
 XX V48830;  
 DT 15-OCT-1998 (first entry)  
 XX ErbB-2 gene antisense oligonucleotide ErbB-2-N-39.  
 DE ErbB-2; antisense oligonucleotide; modulate; gene expression; ss.  
 KW ErbB-2; antisense oligonucleotide; modulate; gene expression; ss.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX EP856579-A1.  
 PN EP856579-A1.  
 XX 05-AUG-1998.  
 PD 05-AUG-1998.  
 XX 31-JAN-1997; 97EP-0101531.  
 PF 31-JAN-1997; 97EP-0101531.  
 XX 31-JAN-1997; 97EP-0101531.  
 PR (BIOG-) BIOGNOSTIK GES BIOMOLEKULARE DIAGNOSTIK.  
 PA Brysch W, Schlingensiepen K;  
 XX WPI; 1998-400910/35.  
 PT Preparation of antisense oligonucleotide(s) which lack long runs of  
 PT consecutive guanosine or inosine - and have specific ratio of  
 PT residues able to form two or three hydrogen bonds, have greater  
 PT activity and reduced toxicity, used therapeutically or to modulate  
 PT growth of cells in culture  
 XX Example 4; Fig 6c; 286pp; English.  
 PS V48709-886 represent antisense oligonucleotides directed against the  
 XX ErbB-2 gene. Of these, only oligonucleotides V48709-91 resulted  
 CC in significant reduction in ErbB-2 protein expression, while  
 CC oligonucleotides V48792-886 had little effect. The oligonucleotides  
 CC exemplify the invention. The specification describes oligonucleotides  
 CC that contain 8-30 nucleotides, which contain at most 8 nucleotides that  
 CC can each form three hydrogen bonds to cytosine; do not contain four  
 CC consecutive nucleotides able to form three H-bonds each to four  
 CC consecutive cytosines; do not contain two sequences of three consecutive  
 CC nucleotides each able to form three H-bonds to three consecutive  
 CC cytosines, and the ratio between residues able to form two H-bonds each  
 CC (2R) or three such bonds (3R) is given by 2R/3R = 0.33-0.72. The  
 CC oligonucleotides are used to modulate expression of genes, particularly  
 CC the genes for p53, ErbB-2, junB, junD, TGF-beta 1 or beta 2 to control  
 CC proliferation of primary cell cultures (e.g. bone marrow stem, liver or  
 CC kidney cells, osteoclasts, osteoblasts and/or keratinocytes). The  
 CC oligonucleotides can also be used to analyse function of proteins (by  
 CC altering their expression or activity) and therapeutically, e.g. in cases  
 CC of cancer or (targeting TGF) for stimulating the immune system.

CC See 078352-078488. The sequences given in GENESQ files  
 CC 078352-078407 and 078488 are antisense oligodeoxynucleotides of  
 CC TGF-beta 1. The sequences given in GENESQ files 078408-78487 are  
 CC antisense oligodeoxynucleotides of TGF-beta 2 in the form of  
 CC phosphorothioate analogues.  
 XX Sequence 14 BP; 3 A; 3 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 15; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
 DB 3 tgttggg 9

RESULT 10  
 V97198/c  
 ID V97198 standard; RNA: 14 BP.  
 XX V97198;  
 AC V97198;  
 XX V97198;  
 DT 01-MAR-1999 (first entry)  
 XX Potato citrate synthase target sequence position 123.  
 DE Solanidine; glucosyltransferase; potato; citrate synthase; target;  
 KW Solanidine; glucosyltransferase; potato; citrate synthase; target;  
 KW hammerhead ribozyme; hairpin ribozyme; alkaloid biosynthesis;  
 KW flower formation; cleavage; solanaceous plant; ss.  
 XX Solanum tuberosum.  
 OS Solanum tuberosum.  
 PN WO9832843-A2.  
 XX 30-JUL-1998.  
 PD 30-JUL-1998.  
 XX 14-JAN-1998; 98WO-US00738.  
 PF 24-NOV-1997; 97US-0979416.  
 PR 28-JAN-1997; 97US-0036545.  
 PR 28-JAN-1997; 97US-0036599.  
 XX (RIBO-) RIBOZYME PHARM INC.  
 PA McSwiggen JA, Zwick MG;  
 XX WPI; 1998-427939/36.  
 PT New enzymatic nucleic acid(s) - useful for, e.g. reducing alkaloid  
 PT biosynthesis or regulating flowering  
 XX Claim 54; Page 59; 79pp; English.  
 PS The present invention describes enzymatic nucleic acid molecules with  
 CC RNA-cleaving activity (e.g. ribozymes) which are capable of modulating  
 CC the expression of plant genes; (i) involved in biosynthesis of  
 CC alkaloids; or (ii) involved in flower formation. V95982 to V96334, and  
 CC V96335 to V96354 represent potato solanidine glucosyltransferase  
 CC hammerhead and hairpin ribozymes, respectively. V95629 to V95981, and  
 CC V96355 to V96734 represent potato solanidine glucosyltransferase target  
 CC sequences. V96773 to V97170, and V97171 to V97195 represent potato  
 CC citrate synthase hammerhead and hairpin ribozymes, respectively. V96735  
 CC to V96772, and V97196 to V97220 represent potato citrate synthase target  
 CC sequences. Ribozymes of the present invention can be used to inhibit  
 CC the synthesis of toxic alkaloids in solanaceous plants, particularly  
 CC potato but also tomato, pepper, aubergine and datura or to inhibit  
 CC flowering in potato, lettuce, spinach, cabbage, brussel sprouts,  
 CC arugula, kale, collards, chard, beet, turnip, sweet potato and turf  
 CC grass. Also the ribozymes can be used for RNA manipulation in the same  
 CC way that restriction endonucleases are for DNA, as well as to examine  
 CC genetic drift and mutations in plants and to detect specific RNA. The

SQ Sequence 14 BP; 5 A; 4 C; 0 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
DB 11 TGTGGA 5

# RESULT 12

V48832/c  
ID V48832 standard; DNA; 14 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

SQ Sequence 14 BP; 5 A; 4 C; 0 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
DB 11 TGTGGA 5

# RESULT 12

V48832/c  
ID V48832 standard; DNA; 14 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

SQ Sequence 14 BP; 5 A; 4 C; 0 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
DB 11 TGTGGA 5

# RESULT 12

V48832/c  
ID V48832 standard; DNA; 14 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

SQ Sequence 14 BP; 5 A; 4 C; 0 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
DB 11 TGTGGA 5

# RESULT 12

V48832/c  
ID V48832 standard; DNA; 14 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

SQ Sequence 14 BP; 5 A; 4 C; 0 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
DB 11 TGTGGA 5

# RESULT 12

V48832/c  
ID V48832 standard; DNA; 14 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX SQ Sequence 14 BP; 5 A; 1 C; 4 G; 4 U; 0 other;

Query Match 100.0%; Score 7; DB 20; Length 14;  
Best Local Similarity 57.1%; Pred. No. 1.4e+04;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgttggga 7  
Db 5 uguugga 11

RESULT 14  
V92044/c  
ID V92044 standard; RNA; 14 BP.  
AC V92044;  
XX  
XX 18-FEB-1999 (first entry)  
XX Human C-raf target sequence nucleotide position 1952.  
DE Human; c-raf; A-raf; B-raf; hammerhead ribozyme; hairpin ribozyme;  
KW target; substrate; catalyst; modulation; expression; Raf gene;  
KW delivery; screening; identification; synthesis; deprotection;  
KW purification; cancer; inflammation; psoriasis; non-hepatic ascites;  
KW infection; genetic drift; restenosis; rheumatoid arthritis; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO9850530-A2.  
XX  
PD 12-NOV-1998.  
XX  
XX 05-MAY-1998; 98WO-US09249.  
XX  
XX 19-DEC-1997; 97US-0068212.  
PR 09-MAY-1997; 97US-0046059.  
PR 09-JUN-1997; 97US-0049002.  
PR 03-JUL-1997; 97US-0051718.  
PR 22-AUG-1997; 97US-0056808.  
PR 02-OCT-1997; 97US-0061321.  
PR 02-OCT-1997; 97US-0061324.  
PR 05-NOV-1997; 97US-0064866.  
XX  
XX (RIBO-) RIBOZYME PHARM INC.  
XX  
XX Beaudry A, Beigelman L, Bellon L, Burgin A, Jarvis T;  
PI Karpeisky A, Kisich K, Matulic-Adamic J, McSwiggen JA;  
PI Parry T, Reynolds M, Sweedler D, Thompson J, Workman CT;  
XX  
XX WPI: 1999-009494/01.  
DR  
XX Identifying new catalytic nucleic acid that modulates selected  
XX processes, especially ribozymes that cleave Raf RNA for treating  
XX cancer, restenosis, and also new ribozymes and modified nucleoside  
XX triphosphates used as antiviral agents and synthons  
XX  
XX Claim 179; Page 156; 259pp; English.  
PS  
XX A method has been developed for the identification of a nucleic acid  
XX capable of modulating a process in a biological system. The method  
XX comprises: (a) introducing into the system a random library of nucleic  
XX acid catalysts (NAC) having a substrate binding domain (SBD), comprising  
XX a random sequence, and a catalytic domain (CD); and (b) identifying NAC  
XX in systems where modulation has occurred and/or determining the sequence  
XX of at least part of the SBDs in such systems. Nucleic acid molecules with  
XX endonuclease activity and catalytic activity, from the present invention,  
XX are used to modulate gene expression in plant and mammalian cells and to  
XX cleave target nucleic acid, particularly for treating systemic diseases  
XX caused by specific RNA, e.g. cancer, inflammation, psoriasis, non-hepatic  
XX ascites and infection. They may also be used to detect genetic drift and

CC mutations in diseased cells and to determine c-raf RNA. Specifically NACs  
CC with RNA-cleaving activity that modulate expression of the Raf gene, are  
CC used to treat cancer, restenosis, psoriasis or rheumatoid arthritis, or  
CC generally any condition associated with the level of c-raf. Introduction  
CC of sugar/phosphate modifications increases stability against nuclease  
CC and activity. V90922 to V93877 represent NACs that can be used in the  
CC method, specifically for modulating the expression of a Raf gene.  
XX  
SQ Sequence 14 BP; 4 A; 5 C; 3 G; 2 U; 0 other;

Query Match 100.0%; Score 7; DB 20; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4e-04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgttggga 7  
Db 14 TGTGGA 8

RESULT 15  
297785/c  
ID 297785 standard; DNA; 14 BP.  
XX  
XX AC 297785;  
XX  
XX 26-APR-2000 (first entry)  
XX  
XX HIV-1 protease gene probe SEQ ID NO:275.  
XX  
XX Human immunodeficiency virus; HIV; protease; probe; detection;  
KW drug selected mutation; hybridisation; genotyping; infection;  
KW drug resistance; ss.  
XX  
XX Human immunodeficiency virus type 1.  
XX  
XX WO9967428-A2.  
XX  
XX 29-DEC-1999.  
XX  
XX 22-JUN-1999; 99WO-EP04317.  
XX  
XX 24-JUN-1998; 98EP-0870143.  
XX  
XX (INNO-) INNOGENETICS NV.  
XX  
XX Stuyver L;  
XX  
XX WPI: 2000-147219/13.  
XX  
XX Detection of drug-selected mutations in the HIV protease gene used to  
XX treat HIV infections  
XX  
XX Claim 3; Page 40; 76pp; English.  
PS  
XX The present invention describes the detection of drug-selected mutations  
XX in the HIV protease gene. The method of detection allows the simultaneous  
XX characterisation of a range of codons involved in drug resistance using  
XX sets of probes optimised to function together in a reverse-hybridisation  
XX assay. 297517 to 297997 represent specifically claimed probes for use in  
XX the assay, and 297479 to 297501 represent specifically claimed HIV  
XX protease gene polymorphic nucleotide sequences. 297502 to 297515, and  
XX 298004 to 298007, represent PCR primers for the HIV protease gene, and  
XX 297516 represents an HIV protease probe used in an example from the  
XX present invention. The method, probes and primers can be used for the  
XX detection of drug-selected mutations in the HIV protease gene. The  
XX method allows the simultaneous characterisation of a range of codons  
XX involved in drug resistance. The method may also be used for HIV  
XX protease genotyping assays. The probes are able to discriminate between  
XX wild type and mutated protease sequences. The method allows rapid and  
XX reliable detection of drug-selected mutation in HIV.  
XX  
XX Sequence 14 BP; 5 A; 6 C; 0 G; 3 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ttttggga 7  
      |||||  
Db 12 TcTtGGGA 6

Search completed: March 27, 2001, 08:22:59  
Job time: 5301 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:17:06 ; Search time 2286.42 Seconds  
(without alignments)  
49.243 Million cell updates/sec

Title: US-09-380-826A-7  
Perfect score: 22  
Sequence: 1 tgttgatcacagatttgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_ph:\*
- 6: gb\_pl1:\*
- 7: gb\_pl2:\*
- 8: gb\_pr1:\*
- 9: gb\_pr2:\*
- 10: gb\_pr3:\*
- 11: gb\_ro:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: em\_fun:\*
- 15: em\_hum1:\*
- 16: em\_hum2:\*
- 17: em\_in:\*
- 18: em\_om:\*
- 19: em\_or:\*
- 20: em\_ov:\*
- 21: em\_pat:\*
- 22: em\_ph:\*
- 23: em\_pl:\*
- 24: em\_ro:\*
- 25: em\_sts:\*
- 26: em\_sy:\*
- 27: em\_un:\*
- 28: em\_vi:\*
- 29: gb\_ba3:\*
- 30: gb\_in1:\*
- 31: gb\_in2:\*
- 32: gb\_in3:\*
- 33: gb\_pl3:\*
- 34: gb\_pr4:\*
- 35: em\_bal:\*
- 36: em\_ba2:\*
- 37: em\_htg1:\*
- 38: em\_htg2:\*
- 39: em\_htg3:\*
- 40: em\_htg4:\*
- 41: em\_htg5:\*
- 42: em\_htg6:\*
- 43: em\_htg7:\*

44: em\_htg8:\*

45: em\_htg9:\*

46: em\_htg10:\*

47: em\_hum3:\*

48: em\_hum4:\*

49: em\_hum5:\*

50: em\_hum6:\*

51: gb\_pr5:\*

52: gb\_pr6:\*

53: gb\_pr7:\*

54: gb\_htg1:\*

55: gb\_htg2:\*

56: gb\_htg3:\*

57: gb\_htg4:\*

58: gb\_htg5:\*

59: gb\_htg6:\*

60: gb\_htg7:\*

61: gb\_htg8:\*

62: gb\_htg9:\*

63: gb\_htg10:\*

64: gb\_htg11:\*

65: gb\_htg12:\*

66: gb\_htg13:\*

67: gb\_htg14:\*

68: gb\_htg15:\*

69: gb\_htg16:\*

70: gb\_htg17:\*

71: gb\_htg18:\*

72: gb\_htg19:\*

73: gb\_htg20:\*

74: gb\_htg21:\*

75: gb\_htg22:\*

76: gb\_htg23:\*

77: gb\_sts1:\*

78: gb\_sts2:\*

79: gb\_vil:\*

80: gb\_vil2:\*

81: gb\_pat1:\*

82: gb\_pat2:\*

83: em\_htg0:\*

84: gb\_htg24:\*

85: gb\_pr8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	22	100.0	1481	2	LFU60594
2	20.4	92.7	354	1	AB007014
3	18.8	85.5	283	1	AF003953
4	18.8	85.5	288	2	LIU94974
5	18.8	85.5	288	2	LIU94975
6	18.8	85.5	288	2	LIU94976
7	18.8	85.5	288	2	LIU94977
8	18.8	85.5	288	2	LIU94978
9	18.8	85.5	288	2	LIU94979
10	18.8	85.5	353	1	AB007012
11	18.8	85.5	1012	7	ATHRS2X
12	18.8	85.5	72590	6	AC003000
13	18.8	85.5	109723	72	AF165146
14	18.8	85.5	200799	6	AF137379
15	18.8	85.5	200799	6	AF137379
16	18.4	83.6	1812	31	AF220067
17	18.4	83.6	27365	58	AC015209
18	18.4	83.6	185087	54	AC007475
19	18.4	83.6	262731	30	AE003823
20	17.8	80.9	1874	79	EHBP44NCP
21	17.8	80.9	31214	33	SPCC830
					U60594 Leptosira
					AB007014 Spirochae
					AF003953 Leptospir
					U94974 Leptosira
					U94975 Leptosira
					U94976 Leptosira
					U94977 Leptosira
					U94978 Leptosira
					U94979 Leptosira
					AB007012 Spirochae
					M98336 Arabidopsis
					AC003000 Arabidops
					AF165146 Homo sapi
					AF137379 Nephrosel
					AF137379 Nephrosel
					AF220067 Drosophil
					AC015209 Drosophil
					AC007475 Drosophil
					AE003823 Drosophil
					L48441 Echinoschloa
					AL109850 S.pombe c

Wed Mar 28 14:03:26 2001

```

AB009052 Arabidops
AC026327 Homo sapi
AC026320 Homo sapi
AC055742 Homo sapi
AC003114 Arabidops
AB008265 Arabidops
AC068144 Homo sapi
AC021015 Homo sapi
AC023155 Homo sapi
AF175672 Unculture
AF2171231 Albugo ca
AF018567 Unidentif
D49367 Lithospermum
Z73136 S.cerevisia
U40832 Strongyloce
U41016 Caenorhabdi
U41345 Caenorhabdi
AL049803 Arabidops
AC083790 Homo sapi
AC036233 Homo sapi
AP002082 Homo sapi
AL078460 Arabidops
Z59052 Human DNA s
10-SEP-1998
partial sequence.

```

# \*ALIGNMENTS

```

RESULT 1
LFU60594 1481 bp DNA BCT 10-SEP-1998
LOCUS Leptospiira fainel 16S ribosomal RNA gene, partial sequence.
DEFINITION U60594
ACCESSION U60594.1 GI:1408219
VERSION
KEYWORDS
SOURCE
ORGANISM
Leptospiira fainel.
Bacteria; Spirochaetales; Leptospiaceae; Leptospiira.
REFERENCE
1 (bases 1 to 1481)
AUTHORS Perolat,P., Chappel,R.J., Adler,B., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F. and Serrano,M.S.
Leptospiira fainel sp. nov., isolated from pigs in Australia
Int. J. Syst. Bacteriol. 48 Pt 3, 851-858 (1998)
98404550
2 (bases 1 to 1481)
AUTHORS Adler,B., Chappel,R.J., Baranton,G., Bulach,D.M.,
Billinghurst,M.L., Letocart,M., Merien,F., Serrano,M.S. and
Perolat,P.
Direct Submission
Submitted (12-JUN-1996) Microbiology, Monash University, Wellington
Rd., Clayton, VIC 3168, Australia
Location/Qualifiers
1. .1481
/organism="Leptospiira fainel"
/strain="Hurstbridge"
/db_xref="taxon:48782"
<1. >1481
/product="16S ribosomal RNA"
BASE COUNT 391 a 335 c 439 g 314 t 2 others
ORIGIN
Query Match 100.0%; Score 22; DB 2; Length 1481;
Best Local Similarity 100.0%; Pred. No. 3.9; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;
QY 1 ttttgatcacaaagatttgata 22
|||||
DB 155 TGTGGATCACAAAGATTGATA 176

```

```

AB007014 354 bp DNA BCT 13-OCT-1997
LOCUS Spirochaeta sp. 16S rRNA gene, partial sequence.
DEFINITION AB007014
ACCESSION AB007014
VERSION 1 GI:2516255
KEYWORDS 16S ribosomal RNA.
SOURCE Spirochaeta sp. (sub_species:Freshwater obligate oligotroph,
strain:FO-95) DNA.
ORGANISM Spirochaeta sp. Spirochaetales; Spirochaetaceae; Spirochaeta.
Bacteria; Spirochaetales; Spirochaetales; Spirochaetales.
REFERENCE
1 (bases 1 to 354)
AUTHORS Shin,M.-S.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-1997) to the DDBJ/EMBL/GenBank databases. Mi-Sun
Shin, Laboratory of Marine Molecular Microbiology, Faculty of
Agriculture, Kyoto University; Oiwa-cho, Kitashirakawa, Sakyo-ku,
Kyoto, Kyoto 606-01, Japan (E-mail:misun@kais.kyoto-u.ac.jp,
Tel:075-753-6224, Fax:075-753-6226)
2 (sites)
AUTHORS Shin,M., Yoshinaga,I., Uchida,A. and Ishida,Y.
TITLE Phylogenetic analysis by 16S rRNA gene sequencing of Obligate
Oligotrophs isolated from the northern basin of Lake Biwa
(Mesotrophic Lake)
JOURNAL Unpublished (1997)
FEATURES
source
1. .354
/organism="Spirochaeta sp."
/strain="FO-95"
/sub_species="Freshwater obligate oligotroph"
/db_xref="taxon:28185"
<1. >354
/product="16S rRNA"
BASE COUNT 91 a 79 c 111 g 73 t
ORIGIN
Query Match 92.7%; Score 20.4; DB 1; Length 354;
Best Local Similarity 95.5%; Pred. No. 25; Indels 0; Gaps 0;
Matches 21; Conservative
QY 1 ttttgatcacaaagatttgata 22
|||||
DB 133 TGTGGATCACAAAGATTGATA 154

```

```

RESULT 3
AF003953 283 bp DNA BCT 30-MAY-1998
LOCUS Leptospiira interrogans strain 48/95 16S ribosomal RNA gene, partial
DEFINITION sequence.
ACCESSION AF003953
VERSION AF003953.1 GI:3169306
KEYWORDS
SOURCE Leptospiira interrogans.
ORGANISM Leptospiira interrogans.
Bacteria; Spirochaetales; Leptospiaceae; Leptospiira.
REFERENCE
1 (bases 1 to 283)
AUTHORS Woo,T.H.S., Patel,B.K.C., Cinco,M., Smythe,L.D., Symonds,M.,
Norris,M. and Dohnt,M.
TITLE Identification of Leptotema by real-time homogeneous assay of rapid
cycle PCR product
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 283)
AUTHORS Woo,T.H.S., Patel,B.K.C., Cinco,M., Smythe,L.D., Symonds,M.,
Norris,M. and Dohnt,M.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1997) School of Science, Griffith University,
Brisbane, QLD 4111, Australia
FEATURES
source
1. .283
/organism="Leptospiira interrogans"
/strain="48/95"
/db_xref="taxon:173"

```





Wed Mar 28 14:03:26 2001

```

VERSION      U94977.1 GI:2735449
KEYWORDS
SOURCE       Leptospiira inadai.
ORGANISM     Bacteria; Spirochaetales; Leptosiraceae; Leptospiira.
REFERENCE    1 (bases 1 to 288)
AUTHORS      Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
              Brenner,D.J. and Patel,B.K.C.
TITLE        Identification of Leptospiira inadai by continuously monitoring
              fluorescence during rapid cycle PCR
JOURNAL       Unpublished
REFERENCE    2 (bases 1 to 288)
AUTHORS      Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
              Brenner,D.J. and Patel,B.K.C.
TITLE        Direct Submission
JOURNAL       Submitted (24-MAR-1997) School of Science, Griffith University,
              Brisbane, QLD 4111, Australia
FEATURES     source
              1..288
              /organism="Leptospiira inadai"
              /strain="268/95"
              /db_xref="taxon:29506"
              <1..>288
              /product="16S ribosomal RNA"
              77 a 61 c 88 g 62 t

rRNA
BASE COUNT   77 a 61 c 88 g 62 t
ORIGIN
Query Match   85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagattgata 22
| | | | | | | | | | | | | | | |
Db 141 TATTGGATCACAGGATTGATA 162

RESULT 8
LOCUS      LIU94978      288 bp      DNA      BCT      01-JAN-1998
DEFINITION Leptospiira inadai 16S ribosomal RNA gene, partial sequence.
ACCESSION  U94978
VERSION     U94978.1 GI:2735450
KEYWORDS
SOURCE      Leptospiira inadai.
ORGANISM     Bacteria; Spirochaetales; Leptosiraceae; Leptospiira.
REFERENCE    1 (bases 1 to 288)
AUTHORS      Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
              Brenner,D.J. and Patel,B.K.C.
TITLE        Identification of Leptospiira inadai by continuously monitoring
              fluorescence during rapid cycle PCR
JOURNAL       Unpublished
REFERENCE    2 (bases 1 to 288)
AUTHORS      Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
              Brenner,D.J. and Patel,B.K.C.
TITLE        Direct Submission
JOURNAL       Submitted (24-MAR-1997) School of Science, Griffith University,
              Brisbane, QLD 4111, Australia
FEATURES     source
              1..288
              /organism="Leptospiira inadai"
              /strain="218/95"
              /db_xref="taxon:29506"
              <1..>288
              /product="16S ribosomal RNA"
              77 a 61 c 88 g 62 t

rRNA
BASE COUNT   77 a 61 c 88 g 62 t
ORIGIN
Query Match   85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagattgata 22
| | | | | | | | | | | | | | | |
Db 141 TATTGGATCACAGGATTGATA 162

RESULT 8
LOCUS      LIU94978      288 bp      DNA      BCT      01-JAN-1998
DEFINITION Leptospiira inadai 16S ribosomal RNA gene, partial sequence.
ACCESSION  U94978
VERSION     U94978.1 GI:2735450
KEYWORDS
SOURCE      Leptospiira inadai.
ORGANISM     Bacteria; Spirochaetales; Leptosiraceae; Leptospiira.
REFERENCE    1 (bases 1 to 288)
AUTHORS      Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
              Brenner,D.J. and Patel,B.K.C.
TITLE        Identification of Leptospiira inadai by continuously monitoring
              fluorescence during rapid cycle PCR
JOURNAL       Unpublished
REFERENCE    2 (bases 1 to 288)
AUTHORS      Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
              Brenner,D.J. and Patel,B.K.C.
TITLE        Direct Submission
JOURNAL       Submitted (24-MAR-1997) School of Science, Griffith University,
              Brisbane, QLD 4111, Australia
FEATURES     source
              1..288
              /organism="Leptospiira inadai"
              /strain="218/95"
              /db_xref="taxon:29506"
              <1..>288
              /product="16S ribosomal RNA"
              77 a 61 c 88 g 62 t

rRNA
BASE COUNT   77 a 61 c 88 g 62 t
ORIGIN
Query Match   85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 tgttgatcacaaagattgata 22
| | | | | | | | | | | | | | | |
Db 141 TATTGGATCACAGGATTGATA 162

RESULT 9
LOCUS      LIU94979      288 bp      DNA      BCT      01-JAN-1998
DEFINITION Leptospiira inadai 16S ribosomal RNA gene, partial sequence.
ACCESSION  U94979
VERSION     U94979.1 GI:2735451
KEYWORDS
SOURCE      Leptospiira inadai.
ORGANISM     Bacteria; Spirochaetales; Leptosiraceae; Leptospiira.
REFERENCE    1 (bases 1 to 288)
AUTHORS      Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
              Brenner,D.J. and Patel,B.K.C.
TITLE        Identification of Leptospiira inadai by continuously monitoring
              fluorescence during rapid cycle PCR
JOURNAL       Unpublished
REFERENCE    2 (bases 1 to 288)
AUTHORS      Woo,T.H.S., Smythe,L.D., Symonds,M., Norris,M., Dohnt,M.,
              Brenner,D.J. and Patel,B.K.C.
TITLE        Direct Submission
JOURNAL       Submitted (24-MAR-1997) School of Science, Griffith University,
              Brisbane, QLD 4111, Australia
FEATURES     source
              1..288
              /organism="Leptospiira inadai"
              /strain="1078 VPI"
              /db_xref="taxon:29506"
              <1..>288
              /product="16S ribosomal RNA"
              77 a 61 c 88 g 62 t

rRNA
BASE COUNT   77 a 61 c 88 g 62 t
ORIGIN
Query Match   85.5%; Score 18.8; DB 2; Length 288;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagattgata 22
| | | | | | | | | | | | | | | |
Db 141 TATTGGATCACAGGATTGATA 162

RESULT 10
LOCUS      AB007012      353 bp      DNA      BCT      13-OCT-1997
DEFINITION Spirochaeta sp. 16S rRNA gene, partial sequence.
ACCESSION  AB007012
VERSION     AB007012.1 GI:2516253
KEYWORDS    16S ribosomal RNA.
SOURCE      Spirochaeta sp. (sub-species: Freshwater obligate oligotroph,
              strain: SO-104) DNA.
ORGANISM     Spirochaeta sp.
              Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.
REFERENCE    1 (bases 1 to 353)
AUTHORS      Shin,M.-S.
TITLE        Direct Submission
JOURNAL       Submitted (05-SEP-1997) to the DDBJ/EMBL/GenBank databases. Mi-Sun
              Shin, Laboratory of Marine Molecular Microbiology, Faculty of
              Agriculture, Kyoto University; Oiwakecho, Kitashirakawa, Sakyo-ku,
              Kyoto, Kyoto 306-01, Japan (E-mail: misun@kais.kyoto-u.ac.jp,
              Tel: 075-753-6224, Fax: 075-753-6226)
              2 (sites)
              Shin,M., Yoshinaga,I., Uchida,A. and Ishida,Y.
              Phylogenetic analysis by 16S rRNA gene sequencing of Obligate
              Oligotrophs isolated from the northern basin of Lake Biwa
              (Mesotrophic Lake)
              Unpublished (1997)

REFERENCE    2 (sites)
AUTHORS      Shin,M., Yoshinaga,I., Uchida,A. and Ishida,Y.
TITLE        Phylogenetic analysis by 16S rRNA gene sequencing of Obligate
              Oligotrophs isolated from the northern basin of Lake Biwa
              (Mesotrophic Lake)
              Unpublished (1997)
JOURNAL

```

```

FEATURES
  source      Location/Qualifiers
 1..353      /organism="Spirochaeta sp."
             /strain="SO-104"
             /sub_species="Freshwater obligate oligotroph"
             /db_xref="taxon:28185"
             <L..>353
             /product="16S rRNA"
BASE COUNT   90 a   80 c   108 g   75 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 1; Length 353;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttgatcacagattgata 22
    ||||| ||||| ||||| |||||
Db 132 TGTGTGATCACAGATTCGATA 153

RESULT 11
ATHRS2X
LOCUS       ATHRS2X 1012 bp mRNA PLN 30-OCT-1994
DEFINITION Arabidopsis thaliana ribonuclease (RNS2) mRNA, complete cds.
ACCESSION   M98336
VERSION     M98336.1 GI:289209
KEYWORDS    Arabidopsis thaliana cDNA to mRNA.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1012)
AUTHORS     Taylor,C.B., Bariola,P.A., delCardayre,S.B., Raines,R.T. and
             Green,P.J.
TITLE       RNS2: a senescence-associated RNase of Arabidopsis that diverged
             from the S-RNases before speciation
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 90 (11), 5118-5122 (1993)
MEDLINE     93281708
FEATURES
  source      Location/Qualifiers
 1..1012      /organism="Arabidopsis thaliana"
             /db_xref="taxon:3702"
 16..795     /gene="RNS2"
 16..795     /gene="RNS2"
             /codon_start=1
             /product="ribonuclease"
             /protein_id="AAA51406.1"
             /db_xref="GI:289210"
             /translation="NASRLCLLLVACIAGFAGDVIELNRSOREDFYFALSLOWPQT
YCRGTRHCCSKNACCRGSDAPQFTIHGLWPDYNDGSMPCCYRSDRDEKEKETSLMDG
LEKYWPSLCGSPSCNGKGSFGWEHEKHGTCSPVPHDEYNVFLTLNLYLKHNV
TDVLYOAGVANSSEKYPVGGIVTAIQNAFHITPEVVCVKRDAIDIRICFCFYKDFKPRD
CVGSDQLTSRKSCPIYVSLPEYTPLDGAMVLMKMPTEAL"
BASE COUNT   287 a   197 c   217 g   311 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 7; Length 1012;
Best Local Similarity 90.9%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgttgatcacagattgata 22
    ||||| ||||| ||||| |||||
Db 672 TGTGTGATCACAGATTGACA 693

RESULT 12
AC003000
LOCUS       AC003000 72590 bp DNA PLN 05-APR-2000

```

```

DEFINITION Arabidopsis thaliana chromosome II section 214 of 255 of the
complete sequence. Sequence from clones T517.
ACCESSION   AC003000 AF002093
VERSION     AC003000.2 GI:6598383
KEYWORDS    HTG
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 72590)
AUTHORS     Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
             Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
             Feldblum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
             Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
             Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Cressy,T.H.,
             Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
             Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
             Venter,J.C.
TITLE       Sequence and analysis of chromosome 2 of the plant Arabidopsis
             thaliana
JOURNAL     Nature 402 (6763), 761-768 (1999)
MEDLINE     20083487
PUBMED      10617197
REFERENCE   2 (bases 1 to 72590)
AUTHORS     Lin,X.
JOURNAL     Direct Submission
COMMENT     Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
             Medical Center Dr., Rockville, MD 20850, USA
             On Dec 17, 1999 this sequence version replaced gi:2642152.
             The sequence and annotation of this chromosome 2 were merged from those
             of the individual clones on this chromosome after removing
             overlaps. For detailed information, please see the TIGR web site
             (http://www.tigr.org/tldb/at/at.html).
             Genes were identified by a combination of three methods: Gene
             prediction programs including GRAIL
             (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
             University of Washington), Genscan (Chris Burge,
             http://genome.stanford.edu/GENSCANW.html), and NetPlantGene
             (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
             complete sequence against a peptide database and plant EST
             databases at TIGR, and manual curations based on those analyses.
             Annotated genes are named to indicate the level of evidence for
             their annotation. Genes with similarity to other proteins are named
             after the database hits. Genes without significant peptide
             similarity but with EST similarity are named as 'unknown' proteins.
             Genes without protein or EST similarity, that are predicted by two
             or more gene prediction programs over most of their length are
             annotated as 'hypothetical' proteins. Genes encoding tRNAs are
             predicted by tRNAscan-SE (Sean Eddy,
             http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
             identified by repeatmasker (Arian Smit,
             http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
             numbered from the top to bottom of the chromosome.
             We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
             F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
             F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
             and Satoshi Tabata for helpful assistance. In addition, we would
             like to thank the TIGR Bioinformatics Department, especially Lixin
             Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
             Peterson, Michael Holmes, and Delwood Richardson for software and
             database support.
             This work was supported by the National Science Foundation,
             Department of Energy and the US Department of Agriculture.
             Address all correspondence to: at@tigr.org.
             Location/Qualifiers
             1..72590
             /organism="Arabidopsis thaliana"
             /cultivar="Columbia"

```

/db\_xref="taxon:3702"  
/chromosome="11"  
1. >72590  
/note="Sequence from clone T517"  
complement(130..198)  
/rpt\_family="(CAAA)n"  
complement(1000..1024)  
/rpt\_family="POLY\_A"  
complement(1266..1291)  
/rpt\_family="(CAAAA)n"  
complement(2759..2862)  
/rpt\_family="POLY\_A"  
complement(<4584..>5909)  
/gene="At2g39710"  
complement(<4584..>5909)  
/note="T517.1: similar to  
GP12245012[gnl|PID|e327000|297341]"  
/gene="At2g39710"  
/note="unknown protein"  
/codon\_start=1  
/protein\_id="AAB87120.1"  
/db\_xref="GI:2642153"  
/translation="MDSSSSLSLSKNFLRISVLLIFPFCSTSTNQLFLSLK  
TKLQSSSDKLSRPHNTLVTLAVGPPONISWLDLTGSELNHLCKSPNLSGVF  
NPVSSITSPVCSPICTRTDLPASPCKDTHLCHVAISYADATSTEGNLAHET  
FVIGSVTRPFLFGCMGSLGNSSEEDAKSTGLCMNRGSLFVNOLGFSKSLSLK  
SDSSGFLIGDASYSWLGPIQITPLVLQSDGPIFDPRVAYIVQLGIRVGSKILSLK  
SVFVDHTGAGTYSWLGPIQITPLVLQSDGPIFDPRVAYIVQLGIRVGSKILSLK  
DLCYKVTRENFSGPLPMVSMFRGAEVSQKLYRVNAGSEKKEVYCYFTFN  
SLLGLTEAPVIGHHQQNMWMEFDLAKSRVGFAGNVRCDLASQRLGRD"  
complement(<10039..11304)  
/gene="At2g39720"  
complement(<10039..11304)  
/note="T517.2"  
/gene="At2g39720"  
complement(10039..11244)  
/note="unknown protein"  
/codon\_start=1  
/protein\_id="AAB87121.1"  
/db\_xref="GI:2642154"  
/translation="MASGSWYCSYSCSFVWVSDTSICPCDGGFLELQEPDLFTPSD  
SFTTQTHRSPTREPPSSSTPSAMHADNSPTPTIVTRSNRSPNVIVLRG  
SAAAPSDVSEGLDSRASFQMYDDGTSGLRPLPSPTELLGSGDFRLLDQISQIE  
LNTNRNLRSCHEHPKSALEALPLIEIDPTLLSDSHCAVCKENFVLSAREMP  
CNHIIYPCDLPLWLNTRNSCPVCRHELPAEDLTDGTAALTAVTAAEEEDSAAAGLT  
INRLPGGGFAGVRIPGMRGGRDMRVYVTEVDGELGLRDLPRVANGSRGRGDDG  
GSRERGFGAGRLMRLFGCGSGSGSIAAAAAASSGSGSRIRVTRTRSFMSFSTASS  
SSRRRNWLA"  
complement(11453..11512)  
/rpt\_family="POLY\_A"  
complement(join(13323..13599,13724..14005,14087..14559,  
14651..14736,14836..14941,15033..15349,15834..16055))  
/gene="At2g39730"  
complement(join(13323..13599,13713..14005,14087..14559,  
14651..14736,14836..14941,15033..15349,15834..16055))  
/gene="At2g39730"  
complement(13323..16055)  
/note="alternative splice form 1 GP1166834|1686720"  
/codon\_start=1  
/product="Rubsico activase"  
/protein\_id="AAB87122.1"  
/db\_xref="GI:2642155"  
/translation="MAAAVSTVGAINRAPLSLNGSGGAVSAPASTFLGKKVTVTSRF

AQSNKNGSFKVLAVREKQDGTGDRWRGLAYDTSDQDITRGKGMVDSVFOAPMGT  
GTHHVLSSSYEQSLRQYLNNDMGFYIAPAFMDGLVHVHITKNFLTNKVPPLI  
LGIMGKGQSGFQCELVMAKMGINPIMMAGELESNGAGEPAKLIIRORYREAAELIK  
KGKWCCLFINDLAGAGRMGTTOYTVNNQMVNATLMNIADNPNTNQLPGMYKAEENA  
RVPIICTGNDSFTLYAPLIRDRMEKFYMAPTREDRIGVCKGIFRTDKIKDEDIVTLV  
DOFGQSIDFGALRARVYDDEVRFVESLGVKIGRLVNSREGPPVFEQPEMTYK  
LMEYGNMLVMEQENVKRVOLAETLYLSQAALGDANADAIGRGTFTYKTEEPFSK"  
CTDPVAENFDPTARSDDGTCVNF"  
complement(join(13579..13599,13713..14005,14087..14559,  
14651..14736,14836..14941,15033..15349,15834..15878))  
/gene="At2g39730"  
/note="alternative splice form 2 GP1166835|1686720"  
/codon\_start=1  
/product="Rubsico activase"  
/protein\_id="AAB87137.1"  
/db\_xref="GI:2642170"  
/translation="MAAAVSTVGAINRAPLSLNGSGGAVSAPASTFLGKKVTVTSRF  
AQSNKNGSFKVLAVREKQDGTGDRWRGLAYDTSDQDITRGKGMVDSVFOAPMGT  
GTHHVLSSSYEQSLRQYLNNDMGFYIAPAFMDGLVHVHITKNFLTNKVPPLI  
LGIMGKGQSGFQCELVMAKMGINPIMMAGELESNGAGEPAKLIIRORYREAAELIK  
KGKWCCLFINDLAGAGRMGTTOYTVNNQMVNATLMNIADNPNTNQLPGMYKAEENA  
RVPIICTGNDSFTLYAPLIRDRMEKFYMAPTREDRIGVCKGIFRTDKIKDEDIVTLV  
DOFGQSIDFGALRARVYDDEVRFVESLGVKIGRLVNSREGPPVFEQPEMTYK  
LMEYGNMLVMEQENVKRVOLAETLYLSQAALGDANADAIGRGTFTYKTEEPFSK"  
complement(17561..17639)  
/rpt\_family="(TA)n"  
/join(<18783..18921,19188..19352,19496..19674,19766..19861,  
19965..20034,20131..20261,20593..>21237)  
/gene="At2g39740"  
complement(<18783..>21237)  
/note="T517.4: predicted by genescan"  
/join(18783..18921,19188..19352,19496..19674,19766..19861,  
19965..20034,20131..20261,20593..21237)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="AAB87123.1"  
/db\_xref="GI:2642156"  
/translation="MSRNPFLDPTLOILQVIKPTRADRTTIVTDLRVQLQSVEC  
LRGATVQPPGFSFVSNLFTFRGDLDDISVDLFGSSILETGKOKOTLLGLHLRALRAG  
LWYKLQVHARVPIKVSQGHQISDISIDNLDGLLSRFLWSEIDGPRDLVL  
LVKEWAKHINDSKTGTNSLSLLVIFHFQTCVPAILPLRVPKPSAVDDLTGV  
RKTAEESIAQVTOITSRRLVSECRNLSIGLTGHQIOESLYRTISLPSQHANGMNV  
RNLHCOARPNQOQNMKOSYNTNPPHPPLTQSRPQONTQNNRNQLQGPVQOG  
VNSRVPITQTOQKSPYKSGNRPLKNTSAGSSQNGHIGKPSGHMNGVNSARPATYNG  
complement(join(21380..21726,21839..21916,22124..22246,  
22328..22585,22834..22993,23212..23322,23481..23751,  
23847..23996,24089..>24814))  
/gene="At2g39750"  
complement(21380..>24814)

Query Match 85.5%; Score 18.8; DB 6; Length 72590;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 tcttgatcaccaagatttgata 22  
||||| |||||||  
Db 35807 TGTTGTTTACACAGATTGACA 35828  
  
RESULT 13  
AF165146  
LOCUS  
DEFINITION  
Homo sapiens chromosome 8 map 8q12-8q13 clone CTA-397H3, WORKING  
DRAFT SEQUENCE, 3 unordered pieces.  
AF165146  
ACCESSION  
AF165146.2 GI:8151945  
VERSION  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
human.  
SOURCE  
Homo sapiens



```

/feature="fts1"
1733..3862
/gene="fts1"
/codon_start=1
/product="putative peptidoglycan synthetase"
/protein_id="AAD54780.1"
/db_xref="GI:5880687"
/translacion="MIYPMKRYRFRRLPKHLQIYYAFLSISMIXIALDPYSKR
PMKMHSGTPOYENERMVMIKLSLATVGLFATRLSGQFNRYTELKSVARQOIGQ
VNDQQRKIILDHGDIADLDPAIDYVHPRMCSISLERIAELSPILDSOYLIN
RLDGGSCICLMDHIDINTSAQIPRLGVDGIELVHPORVYPRKGSFESTLGVDTGEG
YQAGLESSLDKMSQIDQPCWMDHGHNFELRPEKQILFHQESALQITIDCGLOE
KVSQITINAMNRFKAKRIALIEAHSGALNCLKATSPSYDPCNYCPHPRFRFCWPT
DLFEPGTFKFNLAIALENGIFOPTDRILDTIRIGDSYDPCNYCPHPRFRFCWPT
TKTOILQRSSNVGRVWQSLDPAIYHNLIRGLGSHRNDNOTFKMSHDHNSGW
NKLDTSDYATSVVQDODEVDHIEAATSFQGLQALMTPLKLLQIATLANGMAVT
PHLSKIVTLDHFLHOSMNEFSIQWVGQSVLSRQYHAKQPRPYTHDLYLGHVPVP
SLELGFDPKSTIPHTERRRLFSRQTCNVLLGMLQVLDQAQTSRGFLPGYAMAG
KGTGAQKASALGGSTDSVTSFVGIYPAVKPKFTVLIIDEDEDPREFGNTAVDT
QTLISEMIVQEDDPPSYFTVSLFRNM"
3998..5059
/gene="fts2"
/codon_start=1
/product="D1 reaction-center protein of photosystem II"
/protein_id="AAD54781.1"
/db_xref="GI:5880688"
/translacion="MTAILERRESTVWARCDWVTSTENRKYIGWFCVLMPIILLTA
TSVFIIGFAAPVDIDIREPVSGLFNGNIIISGAIIPSSAAGIHFIPIWEAASI
DEWLNGGCVELIVHFLTGACVYMGREWELSLRGMFPAVAYSAFVAATAVFLI
YPIQGSQSDGMLGISTGFNMIVFOAEHNLHMFHMLGVAGVFGGSLFSAMHSL
VYSLLRETTENESANAGYKQCESETYVVAACHYFGRLIFQVASFNNRSLHFFLA
AMPVCIVFTALGVSTMAFNGFNENSVQDSQVRVINTWADIINRANLMEVYHER
NAHNPFLDASVDAVAVQ"
5489..6964
/gene="atpA"
/codon_start=1
/product="CF1 beta subunit of ATP synthase"
/protein_id="AAD54782.1"
/db_xref="GI:5880689"
/translacion="MSQAGVKTQAGKIVQIVGPMVMDVAFQPGSMPTNYNLIIVKGRN
GAGQEVSVNCEVOQLGDLVRAVSNATDGLMRGEVDTDRALSPVPGPTILGRIF
NVLGEPDNGVQGNKTLPIHREAPAFVLDLTKLSTFETGKVVVDLLAPYRGKADY
LFGGAGVKTIVLIMELNNTAKAHGCVSVFGVGGERTREGNDLYMCEKSVINKADY
KSSKVALYQOMNEPPGAMRVGLTALTAETFRDYNODVLLFIDNIERFVQAGSEV
SALLGRMPSAVGIQPLATENGLOERITSTKDGSLTSIOAVTPADDLTPAPATTF
AHLDTATVSLRNLAAGIYPVADPLDSTMTLQPLWILGDDHYGTACRVKQTLQRYKEL
QDITAILGDELSEEDRMIVARAKIERFLSPFFVAEFTGAPGKYVPLASIQGFK
LILSGELDSLPEAFYLVGNIEPAIAKASIAQAAAAAAA"
6977..7378
/gene="atpB"
/codon_start=1
/product="CF1 epsilon subunit of ATP synthase"
/protein_id="AAD54783.1"
/db_xref="GI:5880690"
/translacion="MSLQVRLTPEQVFLNVSADEILPTNTGOMGLVTHNTPLITAI
DIGMPLVRSESTWQSMALLGGLALVKNQVILVNEAELGSDINAEAEETFLAAKEA
LANSKRDKQIENNLAKRVRVQATLVK"
7622..8104
/gene="petB"
/codon_start=1
/product="subunit IV of cytochrome b6/f complex"
/protein_id="AAD54784.1"
/db_xref="GI:5880691"
/translacion="MSVTKKPDLTDPVLRALAKMGHNYGEPANWDLNLYMFPVVI
LGLTSCITGLAVLDPAALGEPANFPATPLPEVFPFQLLRTPVKNLLGLLMA
AVPAGLLTVPFIESINKFQNPFRPVATTVELICTVVAIWLIGIGATLPIDISLTGLF
complement(8132..8779)
/gene="petB"
complement(8132..8779)
/gene="petB"
/codon_start=1
/product="apocytochrome b6 of cytochrome b6/f complex"
/protein_id="AAD54785.1"
/db_xref="GI:5880692"
/translacion="MSKIYDWFEELEIQTADDITSKYVPPHVNIFVCFGIIITCF
LQVATGAMFYIYRPTVEAFASVEYIMTNVNEGLNIRSHRSASMMVMMLLHV
RYLVTGGFKPRELTVTGVILAVITVSGFVSLQMDQVGYWAKYIVTGVPAIPV
ICAPVLELGRSVSVQSTLTFRYSLSLTFVLPLLTAVFMLHFLMIRKQGISGGL"
complement(9045..9431)
/gene="tp12"
complement(9045..9431)
/gene="tp12"
/codon_start=1
/product="ribosomal protein L12"
/protein_id="AAD54786.1"
/db_xref="GI:5880693"
/translacion="NSTITNQIEQLKSMTLLEAAELVSIIEETFGVDASAPAGVM
VAAAPSAFVVEQTEFTVINDVPSAKIAVIVKVRALTGLGLKREKDMIESVPKAI
KEGIGKDEAQQIKQOLEEAGATVTK"
complement(9442..9846)
/gene="rps9"
complement(9442..9846)
/gene="rps9"
/codon_start=1
/product="ribosomal protein S9"
/protein_id="AAD54787.1"
/db_xref="GI:5880694"
/translacion="MDHTSSVGLRKTGRKCAVAOYRIIECTCOLTINORPGVPYLO
NPAYLAAQALDVTETLNTYRTVKTGGGLTGQAEAIKGLARALCSLHKHRAK
KPGYLRNPLRKRRKRYGLKARKAPQFSKR"
complement(9998..11485)
/gene="rpoA"
complement(9998..11485)
/gene="rpoA"
/codon_start=1
/product="alpha subunit of RNA polymerase"
/protein_id="AAD54788.1"
Query Match 85.5%; Score 18.8; DB 6; Length 200799;
Best Local Similarity 90.9%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 tggtagatcacagatttgata 22
Db 134224 TGTGGTCACATGATTGATA 134245
RESULT 15
AF137379/c AF137379 200799 bp DNA circular PLN 14-SEP-1999
LOCUS Nephroselmis olivacea chloroplast DNA, complete genome.
DEFINITION AF137379
ACCESSION AF137379.1 GI:5880684
VERSION
KEYWORDS Nephroselmis olivacea.
SOURCE Chloroplast Nephroselmis olivacea
ORGANISM Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;
Chlorodendraceae; Chlorodendraceae; Nephroselmis.
REFERENCE 1 (bases 1 to 200799)
AUTHORS Turnel, M., Otis, C. and Lemieux, C.
TITLE The complete chloroplast DNA sequence of the green alga
Nephroselmis olivacea: insights into the architecture of ancestral
chloroplast genomes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (18), 10248-10253 (1999)
MEDLINE 99398694
REFERENCE 2 (bases 1 to 200799)
AUTHORS Turnel, M., Otis, C. and Lemieux, C.
```

TITLE Direct Submission  
JOURNAL Submitted (24-MAR-1999) Biochimie, Universite Laval, Pavillon  
Charles-Eugene Marchand, Quebec G1K 7P4, Canada

FEATURES  
Location/Qualifiers  
1. .200799  
/organism="Nephroselmis olivacea"  
/organelle="plastid:chloroplast"  
/db\_xref="taxon:31312"  
/cell\_line="NIES 484"  
misc\_feature  
1. .92126  
/note="Region: large single-copy region (LSC)";  
tRNA  
55. .137  
/gene="trnL(gag)"  
/note="codons recognized: CUY"  
/product="trNA-Leu"  
/anticodon=(pos:88. .90,aa:Leu)  
55. .137  
/gene="trnL(gag)"  
332. .634  
/gene="trnL(gag)"  
332. .634  
/gene="trnL(gag)"  
/codon\_start=1  
/product="ribosomal protein S14"  
/protein\_id="AAD54778.1"  
/db\_xref="GI:5880685"  
tRNA  
/translation="WAKKSMIERESKRAALVAKYATRRQALKAAIOKTKSFDRLVLO  
HOLODPVNSVPCRLHNRCTITGRPKGYRDFGLSRHELRAHAGCLLPVTKASH"  
765. .849  
/gene="trnL(uua)"  
/note="codon recognized: UUA"  
/product="trNA-Leu"  
/anticodon=(pos:798. .800,aa:Leu)  
765. .849  
/gene="trnL(uua)"  
983. .1054  
/gene="trnQ(uug)"  
/note="codons recognized: CAR"  
/product="trNA-Gln"  
/anticodon=(pos:1015. .1017,aa:Gln)  
983. .1054  
/gene="trnQ(uug)"  
1102. .1182  
/gene="trnL(uag)"  
/note="codons recognized: CUR"  
/product="trNA-Leu"  
/anticodon=(pos:1136. .1138,aa:Leu)  
1102. .1182  
/gene="trnL(uag)"  
complement(1364. .1468)  
/gene="psbM"  
complement(1364. .1468)  
/gene="psbM"  
/codon\_start=1  
/product="M protein of photosystem II"  
/protein\_id="AAD54779.1"  
/db\_xref="GI:5880686"  
tRNA  
/translation="MEVNILGLIATALFTIIPTSFLILLYVKTASQON"  
1618. .1691  
/gene="trnM(cau)"  
/note="codon recognized: AUG"  
/product="trNA-Met"  
/anticodon=(pos:1652. .1654,aa:Met)  
1618. .1691  
/gene="trnM(cau)"  
1733. .3862  
/gene="ftsI"  
1733. .3862  
/gene="ftsI"  
/codon\_start=1  
/product="putative peptidoglycan synthetase"  
/protein\_id="AAD54780.1"  
/db\_xref="GI:5880687"

gene  
CDS  
/translation="MIYNYPMKYRRQFRRLPKHLOQIYYAFISLSITMWIKIALDPYSKR  
PKMWHSGTPFOYENERMVMIKLSLATVGLLFATRLSGLOFNKYTELKSAVERQQICQ  
VNDPQOKIILDHGDIIVADLPAYDLYVHPRMCSISLERIAELLSPILDSQOYLIN  
RLDEGSGICLMHQIDTNTSAOTRRGLVDGIELVHHQPVRYPKRGSFESILGVYDTEG  
YGAAGLESSLDDMMKSTYODPCWMDGHGFLGIRPKOILFHOESALOLTIDGLOE  
KYSOLITNAMNREGAKRIAAIIMEHSGAIRCLATSPSYDPCYGFWEFRFCWPIT  
DLPEFGSTFKPVLNLALENGEFOPTDRILDYTKIRIGDSWQIGNVGGGFIDWRLDHL  
TGTOILQRSSNVMVRVMSQSLPAIYHRLNIRLIGLSHRNDNQTSPKMSHDHNSGW  
NLKDLTSDYAISSVVDQDEFVDHEIATASFQGLAMTPLKLLQIATITANGMAVT  
PHILISKIVTDHFEHLHLOSNEFSLOGVQSLSRQYHAKOPRPRYTHDLYLGHVVP  
SLELGFWDVKSIPIPHTRERRRLEFSQTCNVLLGMLQVLDQAQATSGRGLPCYANAG  
KTCQAQASALGGYSTDSVTSEVGLYPAVKPKFVTLVILDEPEDPFRGFTAVDVT  
QTLISEMIVQEQDPPSYPTVSLFRNM"  
3998. .5059  
/gene="psbA"  
3998. .5059  
/gene="psbA"  
/codon\_start=1  
/product="D1- reaction-center protein of photosystem II"  
/protein\_id="AAD54781.1"  
/db\_xref="GI:5880688"  
/translation="MTAILERRESTSWARFCDWVTSTENRLYIGWGLMIPLLLLTA  
TSVFIIGFTAAPVDIDGIREPVSGSLFGNNITSCAIIPSSAAIGIHFYPIWEAASI  
DEWLYNGCYELIVLHFLGVACVYCREHELSFRLCMRPWIAYASAPAAATAVFLI  
YPIQGSFSDGHPGLGSGTFNEMIVFOACHNLMHPFRLGVAGVFGSLFSAMRSL  
VTSSLIRETTENESANAGYFGQEEETYNIVAAGYFGRLLIFOYASFNRSRLHFFLA  
AMPVYVIMFTALGVSTMAFNGLNFQNSVDSQGRVINTWADIINRANLGMVMEVHER  
NAHNFPDLIASVADPAVQC"  
5489. .6964  
/gene="atpB"  
5489. .6964  
/gene="atpB"  
/codon\_start=1  
/product="CF1 beta subunit of ATP synthase"  
/protein\_id="AAD54782.1"  
/db\_xref="GI:5880689"  
/translation="MSQAGVKTQAKIVQIVGVPMVDVAFQPGSMNINNALIVKGRN  
GAGQEVSVCEVQQLIGDLGLVRAVSMATDGLMRGMEVTDGHALSVPGVPTGLGRIF  
NVLGEVDNMGPVNEKTLPIHREAPFVLDLTKLSIFETKVVLLAPYRRGGVIG  
LFGGAGVGTGLIMELINNIKAHGVSVFGVGGERTREGNDLYMCEKSKVINKADY  
KSKVALVYQNMNPPGARVRLTALTMAEYFRDYNODVLLFIQNIREFVQAGSEV  
SALLGRMPSAVGYOPTLATFMGLOERITSTKDGSI'SIOAVYVPADDLTDPAPTFF  
AHLDTATVLSRLAAKGIYPAVDPLDSTSTMLOPWLIGDDHYGTAQRVKOTLQRYKEL  
QDITAILGLDELSEEDRMIVARARKTERFELSQPFVFAEVTGAPGYVPLAESIQGFK  
LILSGELSLPESAFYLVGNIETATAKASIAQAAAAA"  
6977. .7378  
/gene="atpE"  
6977. .7378  
/gene="atpE"  
/codon\_start=1  
/product="CF1 epsilon subunit of ATP synthase"  
/protein\_id="AAD54783.1"  
/db\_xref="GI:5880690"  
/translation="MSLOVRILTPEOVFLNVSADETILPTNTGOMGLNHTPLTIAI  
DIGPMLVRSESTWOSMALLGLIAYKDNQVILVNEAELGSDINAEAEATTFLAAEA  
LANSKTRKQDIENNLAFAKRVRYQATLVK"  
complement(7622. .8104)  
/gene="petD"  
complement(7622. .8104)  
/gene="petD"  
/codon\_start=1  
/product="subunit IV of cytochrome b6/f complex"  
/protein\_id="AAD54784.1"  
/db\_xref="GI:5880691"  
/translation="MSVTRKPDLTDPVLRAKLAKGMHNVYGEPAWPNDDLVMFPVVI  
LGTLSCTITGLAVLDPAALIGEPAWNPATPLEIPWYFFPVFOLLTRVKNLLGVLLMA  
AVPAGLLTVPFIESINKFQNPFRPVATVTVLIGTVVAIWLGIGATLPIDISLTGLF  
"  
complement(8132. .8779)  
/gene="petB"  
complement(8132. .8779)  
/gene="petB"  
/codon\_start=1

```
/product="apocytochrome b6 of cytochrome b6/f complex"
/protein_id="AAD54785.1"
/db_xref="GI:5880692"
/translation="MSKIYDFEERLEIQAIADDTISKYVPPHVNIEYCEGGITLTCF
LIQVATGFMTFYRTYFAFASVEYIMTNVFNGLIRSIHRWSASMMVMILLHVF
RVYLTGGFKRPRLTWTGTVILAVITVSGVTGSLPMDQGVWAVKIVTVGVPDAIPV
IGAPLVELLRGSVSGQSTLTRYSLHTFVPLLTAVFMLHFLMIRKQISGPL"
complement(9045..9431)
/gene="rpl12"
CDS
complement(9045..9431)
/gene="rpl12"
/codon_start=1
/product="ribosomal protein L12"
/protein_id="AAD54786.1"
/db_xref="GI:5880693"
/translation="MSITNQIIIEQLKSMTLLAEALVSOIETFGVDASAPAAGAVM
VAAAPSPAVVEEQTEFTVMINDVPSAKRIAVIKVRALIGLGLKEAKDMIESVPKAI
REGICKDPAQOIKOOLEEAGATVTK"
complement(9442..9846)
/gene="rps9"
CDS
complement(9442..9846)
/gene="rps9"
/codon_start=1
/product="ribosomal protein S9"
/protein_id="AAD54787.1"
/db_xref="GI:5880694"
/translation="MDHTSSVGLRKTGRKCAVAQVRIEGTGOLTINQREGVPYLYQ
NPAYLLAAGALDVTETLNRYDTIVKTKGGTLTGQAEAIKGLARALCSLHKKKAL
KPGYLTRNPLRKRRKYGLKARKAPQFSKR"
complement(9998..11485)
/gene="rpoA"
CDS
complement(9998..11485)
/gene="rpoA"
/codon_start=1
/product="alpha subunit of RNA polymerase"
/protein_id="AAD54788.1"
```

```
Query Match      85.5%  Score 18.8; DB 6; Length 200799;
Best Local Similarity 90.9%  Pred. No. 77;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  tgttgatcacagaatttgata 22
        ||||| ||||| ||||| |||||
Db 158702 TGTGGCTCACATGATTGTGATA 158681
```

Search completed: March 27, 2001, 08:17:15  
Job time: 5912 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:17:04 ; Search time 2286.42 Seconds  
(without alignments)  
15.668 Million cell updates/sec

Title: US-09-380-826A-5  
Perfect score: 7  
Sequence: 1 ttgata 7

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_ph: \*  
6: gb\_pl1: \*  
7: gb\_pl2: \*  
8: gb\_pr1: \*  
9: gb\_pr2: \*  
10: gb\_pr3: \*  
11: gb\_ro: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: em\_fun: \*  
15: em\_hum1: \*  
16: em\_hum2: \*  
17: em\_in: \*  
18: em\_om: \*  
19: em\_or: \*  
20: em\_ov: \*  
21: em\_pat: \*  
22: em\_ph: \*  
23: em\_pl: \*  
24: em\_ro: \*  
25: em\_sts: \*  
26: em\_sy: \*  
27: em\_un: \*  
28: em\_v1: \*  
29: gb\_ba3: \*  
30: gb\_in1: \*  
31: gb\_in2: \*  
32: gb\_in3: \*  
33: gb\_pl3: \*  
34: gb\_pr4: \*  
35: em\_ba1: \*  
36: em\_ba2: \*  
37: em\_htg1: \*  
38: em\_htg2: \*  
39: em\_htg3: \*  
40: em\_htg4: \*  
41: em\_htg5: \*  
42: em\_htg6: \*  
43: em\_htg7: \*

44: em\_htg8: \*  
45: em\_htg9: \*  
46: em\_htg10: \*  
47: em\_hum3: \*  
48: em\_hum4: \*  
49: em\_hum5: \*  
50: em\_hum6: \*  
51: gb\_pr5: \*  
52: gb\_pr6: \*  
53: gb\_pr7: \*  
54: gb\_htg1: \*  
55: gb\_htg2: \*  
56: gb\_htg3: \*  
57: gb\_htg4: \*  
58: gb\_htg5: \*  
59: gb\_htg6: \*  
60: gb\_htg7: \*  
61: gb\_htg8: \*  
62: gb\_htg9: \*  
63: gb\_htg10: \*  
64: gb\_htg11: \*  
65: gb\_htg12: \*  
66: gb\_htg13: \*  
67: gb\_htg14: \*  
68: gb\_htg15: \*  
69: gb\_htg16: \*  
70: gb\_htg17: \*  
71: gb\_htg18: \*  
72: gb\_htg19: \*  
73: gb\_htg20: \*  
74: gb\_htg21: \*  
75: gb\_htg22: \*  
76: gb\_htg23: \*  
77: gb\_sts1: \*  
78: gb\_sts2: \*  
79: gb\_vil: \*  
80: gb\_vil2: \*  
81: gb\_pat1: \*  
82: gb\_pat2: \*  
83: em\_htg0: \*  
84: gb\_htg24: \*  
85: gb\_pr8: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	7	100.0	11	81 I20741	I20741 Sequence 6
C 2	7	100.0	11	81 I20742	I20742 Sequence 7
C 3	7	100.0	13	81 AR082373	AR082373 Sequence
C 4	7	100.0	13	81 I52063	I52063 Sequence 5
C 5	7	100.0	13	82 I78419	I78419 Sequence 21
C 6	7	100.0	14	79 AX007639	AX007639 Sequence
C 7	7	100.0	14	81 A88596	A88596 Sequence 74
C 8	7	100.0	14	81 A90563	A90563 Sequence 74
C 9	7	100.0	14	81 AR082368	AR082368 Sequence
C 10	7	100.0	14	82 I78414	I78414 Sequence 21
C 11	7	100.0	15	79 AX007638	AX007638 Sequence
C 12	7	100.0	15	79 AX007675	AX007675 Sequence
C 13	7	100.0	15	81 A35660	A35660 Synthetic h
C 14	7	100.0	15	81 AR045297	AR045297 Sequence
C 15	7	100.0	15	81 AR071525	AR071525 Sequence
C 16	7	100.0	15	81 I52349	I52349 Sequence 90
C 17	7	100.0	15	81 I61575	I61575 Sequence 12
C 18	7	100.0	15	81 I61576	I61576 Sequence 13
C 19	7	100.0	15	81 I61577	I61577 Sequence 13
C 20	7	100.0	15	82 I77301	I77301 Sequence 8
C 21	7	100.0	15	82 I77302	I77302 Sequence 9

[illegible]

QY 1 ttgtata 7  
 |||||  
 Db 10 TTTGATA 4

RESULT 5  
 I78419/c 13 bp DNA PAT 03-APR-1998  
 LOCUS Sequence 217 from patent US 5693535.  
 DEFINITION I78419  
 ACCESSION I78419  
 VERSION I78419.1 GI:3014573  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 13)  
 AUTHORS Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and Thompson,J.D.  
 TITLE HIV targeted ribozymes  
 JOURNAL Patent: US 5693535-A 217 02-DEC-1997;  
 FEATURES Location/Qualifiers  
 source 1..13  
 /organism="unknown"  
 BASE COUNT 5 a 4 c 1 g 3 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 82; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7  
 |||||  
 Db 10 TTTGATA 4

RESULT 6  
 AX007639/c 14 bp DNA VRL 06-SEP-2000  
 LOCUS Sequence 181 from Patent WO9967428.  
 DEFINITION AX007639  
 ACCESSION AX007639  
 VERSION AX007639.1 GI:9995336  
 KEYWORDS Aids-associated retrovirus.  
 SOURCE Aids-associated retrovirus.  
 ORGANISM Viruses; Retroid viruses; Retroviridae.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Stuyver,L.  
 TITLE Method for detection of drug-selected mutations in the hiv protease gene  
 JOURNAL Patent: WO 9967428-A 29-DEC-1999;  
 INNOCENETICS NV (BE); STUYVER LIEVEN (BE)  
 FEATURES Location/Qualifiers  
 source 1..14  
 /organism="Aids-associated retrovirus"  
 /db\_xref="taxon:11966"  
 BASE COUNT 5 a 1 c 2 g 6 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 79; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7  
 |||||  
 Db 11 TTTGATA 5

RESULT 7  
 A88596 14 bp DNA PAT 22-JAN-2000  
 LOCUS Sequence 212 from patent US 5972704.  
 DEFINITION A88596  
 ACCESSION A88596  
 VERSION A88596.1 GI:10009094  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and

DEFINITION Sequence 744 from Patent WO9833904.  
 A88596  
 VERSION A88596.1 GI:6737166  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Brysch,W. and Schlingsenslepen,K.  
 TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD  
 JOURNAL Patent: WO 9833904-A 06-AUG-1998;  
 BIOGNOSTIK GES (DE); BRYSCH WOLFGANG (DE)  
 FEATURES Location/Qualifiers  
 source 1..14  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 5 a 1 c 1 g 7 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7  
 |||||  
 Db 1 TTTGATA 7

RESULT 8  
 A90563 14 bp DNA PAT 22-JAN-2000  
 LOCUS Sequence 744 from Patent EP0856579.  
 DEFINITION A90563  
 ACCESSION A90563.1 GI:6739077  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Brysch,W.D. and Schlingsenslepen,K.D.  
 TITLE An antisense oligonucleotide preparation method  
 JOURNAL Patent: EP 0856579-A 05-AUG-1998;  
 BIOGNOSTIK GES (DE)  
 FEATURES Location/Qualifiers  
 source 1..14  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 5 a 1 c 1 g 7 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7  
 |||||  
 Db 1 TTTGATA 7

RESULT 9  
 A8082368 14 bp DNA PAT 31-AUG-2000  
 LOCUS Sequence 212 from patent US 5972704.  
 DEFINITION A8082368  
 ACCESSION A8082368  
 VERSION A8082368.1 GI:10009094  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and

DEFINITION Sequence 744 from Patent WO9833904.  
 A88596  
 VERSION A88596.1 GI:6737166  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Brysch,W. and Schlingsenslepen,K.  
 TITLE AN ANTISENSE OLIGONUCLEOTIDE PREPARATION METHOD  
 JOURNAL Patent: WO 9833904-A 06-AUG-1998;  
 BIOGNOSTIK GES (DE); BRYSCH WOLFGANG (DE)  
 FEATURES Location/Qualifiers  
 source 1..14  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 5 a 1 c 1 g 7 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtata 7  
 |||||  
 Db 1 TTTGATA 7

RESULT 9  
 A8082368 14 bp DNA PAT 31-AUG-2000  
 LOCUS Sequence 212 from patent US 5972704.  
 DEFINITION A8082368  
 ACCESSION A8082368  
 VERSION A8082368.1 GI:10009094  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 14)  
 AUTHORS Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and

```

Thompson, J.D.
HIV nef targeted ribozymes
JOURNAL Patent: US 5972704-A 212 26-OCT-1999;
FEATURES Location/Qualifiers
    source 1..14
    /organism="unknown"
BASE COUNT 5 a 1 c 4 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 14;
Best Local Similarity 100.0%; Pred. No. 6.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7
    |||||
Db 3 TTTGATA 9

RESULT 10
LOCUS I78414 178414 14 bp DNA PAT 03-APR-1998
DEFINITION Sequence 212 from patent US 5693535.
ACCESSION I78414
VERSION I78414.1 GI:3014568
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Draper, K.G., Chowrira, B., McSwiggen, J., Stinchcomb, D.T. and Thompson, J.D.
TITLE HIV targeted ribozymes
JOURNAL Patent: US 5693535-A 212 02-DEC-1997;
FEATURES Location/Qualifiers
    source 1..14
    /organism="unknown"
BASE COUNT 5 a 1 c 4 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 82; Length 14;
Best Local Similarity 100.0%; Pred. No. 6.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata.7
    |||||
Db 3 TTTGATA 9

RESULT 11
LOCUS AX007638/c 15 bp DNA VRL 06-SEP-2000
DEFINITION Sequence 180 from Patent WO9967428.
ACCESSION AX007638
VERSION AX007638.1 GI:9995335
KEYWORDS
SOURCE Aids-associated retrovirus.
ORGANISM Aids-associated retrovirus
REFERENCE 1 (bases 1 to 15)
AUTHORS Stuyver, L.
TITLE Method for detection of drug-selected mutations in the hiv protease gene
JOURNAL Patent: WO 9967428-A 29-DEC-1999;
FEATURES Location/Qualifiers
    source 1..15
    /organism="Aids-associated retrovirus"
BASE COUNT 5 a 1 c 3 g 6 t
ORIGIN

Query Match 100.0%; Score 7; DB 79; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7
    |||||
Db 12 TTTGATA 6

RESULT 12
LOCUS AX007675/c 15 bp DNA VRL 06-SEP-2000
DEFINITION Sequence 217 from Patent WO9967428.
ACCESSION AX007675
VERSION AX007675.1 GI:9995372
KEYWORDS
SOURCE Aids-associated retrovirus.
ORGANISM Aids-associated retrovirus
REFERENCE 1 (bases 1 to 15)
AUTHORS Stuyver, L.
TITLE Method for detection of drug-selected mutations in the hiv protease gene
JOURNAL Patent: WO 9967428-A 29-DEC-1999;
FEATURES Location/Qualifiers
    source 1..15
    /organism="Aids-associated retrovirus"
BASE COUNT 5 a 1 c 4 g 5 t
ORIGIN

Query Match 100.0%; Score 7; DB 79; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7
    |||||
Db 12 TTTGATA 6

RESULT 13
LOCUS A35660/c 15 bp DNA PAT 02-DEC-1996
DEFINITION Synthetic human IFN-alpha 2 gene oligo.
ACCESSION A35660
VERSION A35660.1 GI:1927042
KEYWORDS
SOURCE Synthetic construct.
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 15)
AUTHORS Camble, R. and Edge, M.D.
TITLE Analogous interferon polypeptides, process for their preparation and pharmaceutical compositions containing them
JOURNAL Patent: EP 0194006-A 105 10-SEP-1986;
FEATURES Location/Qualifiers
    source 1..15
    /organism="synthetic construct"
BASE COUNT 5 a 4 c 2 g 4 t
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7
    |||||

```

Db 12 TTTGATA 6

## RESULT 14

AR045297/C  
LOCUS AR045297 15 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 90 from patent US 5817796.  
ACCESSION AR045297  
VERSION AR045297.1 GI:5966762  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
1 (bases 1 to 15)  
Unclassified.  
AUTHORS Stinchcomb,D.T., Draper,K., McSwiggen,J. and Jarvis,T.  
TITLE C-myb ribozymes having 2'-5'-linked adenylylate residues  
JOURNAL Patent: US 5817796-A 90 06-OCT-1998;  
FEATURES  
Source  
1..15  
Location/Qualifiers  
/organism="unknown"  
BASE COUNT 6 a 3 c 3 g 3 t  
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
|||||||  
Db 11 TTTGATA 5

## RESULT 15

AR071525/C  
LOCUS AR071525 15 bp DNA PAT 18-FEB-2000  
DEFINITION Sequence 25 from patent US 5911982.  
ACCESSION AR071525  
VERSION AR071525.1 GI:7222413  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
1 (bases 1 to 15)  
Unclassified.  
AUTHORS Chao,Y.  
TITLE Hz-1 virus persistence-associated-gene 1 (PAG1) promoter uses  
therefor, and compositions containing same or products therefrom  
JOURNAL Patent: US 5911982-A 25 15-JUN-1999;  
FEATURES  
Source  
1..15  
Location/Qualifiers  
/organism="unknown"  
BASE COUNT 9 a 2 c 2 g 2 t  
ORIGIN

Query Match 100.0%; Score 7; DB 81; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.1e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
|||||||  
Db 12 TTTGATA 6

Search completed: March 27, 2001, 08:17:04  
Job time: 5901 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:20 ; Search time 2517.78 Seconds  
(without alignments)  
19.482 Million cell updates/sec

Title: US-09-380-826A-4  
Perfect score: 7  
Sequence: 1 tgttga 7

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_estl:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthuml:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estrol:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est58:\*  
105: gb\_est59:\*  
106: gb\_est60:\*  
107: gb\_est61:\*  
108: gb\_est62:\*  
109: gb\_est63:\*  
110: gb\_est64:\*  
111: gb\_est65:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

190: gb\_gss25:.\*  
191: gb\_gss26:.\*  
192: gb\_gss27:.\*  
193: gb\_gss28:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7	100.0	19	171	AZ386406	AM0145C22
2	7	100.0	20	173	AZ484701	AM0311C24
3	7	100.0	21	171	AZ361153	AM0104A17
4	7	100.0	21	172	AZ450829	AM0249E13
5	7	100.0	21	174	AZ508369	AM0350A24
6	7	100.0	25	15	AI074857	0427b11.s
7	7	100.0	25	16	AI158948	AI42807.x
8	7	100.0	25	20	AI453394	tj37802.x
9	7	100.0	26	171	AZ352012	AM0090M13
10	7	100.0	29	170	AZ332711	AM0061N14
11	7	100.0	29	172	AZ419519	AM0196N03
12	7	100.0	31	170	AZ333172	AM0062B09
13	7	100.0	31	171	AZ357647	AM0099D19
14	7	100.0	32	170	AZ320254	AM0040P07
15	7	100.0	32	170	AZ328463	AM0052D18
16	7	100.0	32	173	AZ467843	AM0279F06
17	7	100.0	33	141	H48291	H48291.yq77a01.r
18	7	100.0	34	172	AZ408169	AM0179C17
19	7	100.0	35	170	AZ320436	AM0040C03
20	7	100.0	35	170	AZ322327	AM0043P14
21	7	100.0	36	148	AQ026010	EP(2)0461
22	7	100.0	36	170	AZ328537	AM0052B23
23	7	100.0	36	172	AZ433353	AM0219J12
24	7	100.0	37	1	AA011782	AA011782.mg92b03.r
25	7	100.0	37	10	AA636499	VR18f02.r
26	7	100.0	37	140	D20685	HUMGS01661
27	7	100.0	37	145	W70764	yd14e02.r1
28	7	100.0	37	148	AQ025921	1(2)k1010
29	7	100.0	38	113	HSM007685	AL042835.Homo.sapi
30	7	100.0	38	173	AZ495773	AM0331O13
31	7	100.0	39	14	AI001077	AM0331O13
32	7	100.0	39	173	AZ493504	AM0328P12
33	7	100.0	40	13	AA871171	AZ493504
34	7	100.0	40	22	AI588805	AA871171.vq32a05.r
35	7	100.0	40	28	AU009892	AI588805.fb99f01.y
36	7	100.0	40	28	AU009893	AU009892
37	7	100.0	40	28	AU009894	AU009893
38	7	100.0	40	28	AU009895	AU009894
39	7	100.0	40	28	AU010246	AU009895
40	7	100.0	40	28	AU010247	AU010246
41	7	100.0	40	172	AZ447227	AU010247
42	7	100.0	40	173	AZ458268	AZ447227
43	7	100.0	41	140	D25852	AZ458268
44	7	100.0	41	170	AZ324505	D25852.HUMGS04228
45	7	100.0	41	171	AZ361377	AZ324505

ALIGNMENTS

RESULT 1  
AZ386406  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AZ386406 19 bp DNA  
1M0145C22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0145C22 F, DNA sequence.

GSS 02-OCT-2000

AZ386406  
GI:10500106  
house mouse.

117: em\_estp16:.\*  
118: em\_estp17:.\*  
119: em\_estp18:.\*  
120: em\_estp19:.\*  
121: em\_estp20:.\*  
122: em\_estp21:.\*  
123: em\_estp22:.\*  
124: em\_estp23:.\*  
125: em\_estp24:.\*  
126: em\_estp25:.\*  
127: em\_estp26:.\*  
128: em\_estp27:.\*  
129: em\_estp28:.\*  
130: em\_estp29:.\*  
131: em\_estp30:.\*  
132: em\_estp31:.\*  
133: em\_estp32:.\*  
134: em\_estp33:.\*  
135: em\_estp34:.\*  
136: em\_estp35:.\*  
137: em\_estp36:.\*  
138: em\_estp37:.\*  
139: em\_estp38:.\*  
140: em\_estp39:.\*  
141: em\_estp40:.\*  
142: em\_estp41:.\*  
143: em\_estp42:.\*  
144: em\_estp43:.\*  
145: em\_estp44:.\*  
146: em\_estp45:.\*  
147: em\_estp46:.\*  
148: em\_estp47:.\*  
149: em\_estp48:.\*  
150: em\_estp49:.\*  
151: em\_estp50:.\*  
152: em\_estp51:.\*  
153: em\_estp52:.\*  
154: em\_estp53:.\*  
155: em\_estp54:.\*  
156: em\_estp55:.\*  
157: em\_estp56:.\*  
158: em\_estp57:.\*  
159: em\_estp58:.\*  
160: em\_estp59:.\*  
161: em\_estp60:.\*  
162: em\_estp61:.\*  
163: em\_estp62:.\*  
164: em\_estp63:.\*  
165: em\_estp64:.\*  
166: em\_estp65:.\*  
167: em\_estp66:.\*  
168: em\_estp67:.\*  
169: em\_estp68:.\*  
170: em\_estp69:.\*  
171: em\_estp70:.\*  
172: em\_estp71:.\*  
173: em\_estp72:.\*  
174: em\_estp73:.\*  
175: em\_estp74:.\*  
176: em\_estp75:.\*  
177: em\_estp76:.\*  
178: em\_estp77:.\*  
179: em\_estp78:.\*  
180: em\_estp79:.\*  
181: em\_estp80:.\*  
182: em\_estp81:.\*  
183: em\_estp82:.\*  
184: em\_estp83:.\*  
185: em\_estp84:.\*  
186: em\_estp85:.\*  
187: em\_estp86:.\*  
188: em\_estp87:.\*  
189: em\_estp88:.\*



```

ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                1 (bases 1 to 19)
                Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
                Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
                ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
                and Wright,D., Weiss,R.
TITLE          Mouse whole genome scaffolding with paired end reads from 10kb
                plasmid inserts
JOURNAL        Unpublished (2000)
COMMENT        Contact: Robert B. Weiss
                University of Utah Genome Center
                University of Utah
                Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
                84112, USA
                Tel.: 801 585 5606
                Fax: 801 585 7177
                Email: ddunn@genetics.utah.edu
                Insert Length: 10000 Std Error: 0.00
                Plate: 0145 row: C column: 22
                Seq primer: CGTTGTAACGACGCGCCAGT
                Class: plasmid ends
                High quality sequence stop: 19.
FEATURES       Location/Qualifiers
                1..19
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUC1M0145C22"
                /clone_lib="Mouse 10kb plasmid UUC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                /note="vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptored DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pWD42 (gil4732114[gb|AF129072.1]), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptored mouse DNA was annealed to
                adaptored vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
BASE COUNT      5 a      2 c      6 g      6 t
ORIGIN
Query Match      100.0%; Score 7; DB 171; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tggtaga 7
    |||||
Db 11 TGTGGA 17

RESULT 2
LOCUS      AZ484701      20 bp      DNA      GSS      05-OCT-2000
DEFINITION  IM0311C24F Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUC1M0311C24 F, DNA sequence.
ACCESSION  AZ484701
VERSION    AZ484701.1 GI:10649799
KEYWORDS   GSS.

```

```

SOURCE          house mouse.
ORGANISM        Mus musculus
REFERENCE        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                1 (bases 1 to 20)
                Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
                Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
                ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
                and Wright,D., Weiss,R.
TITLE            Mouse whole genome scaffolding with paired end reads from 10kb
                plasmid inserts
JOURNAL          Unpublished (2000)
COMMENT          Contact: Robert B. Weiss
                University of Utah Genome Center
                University of Utah
                Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
                84112, USA
                Tel: 801 585 5606
                Fax: 801 585 7177
                Email: ddunn@genetics.utah.edu
                Insert Length: 10000 Std Error: 0.00
                Plate: 0311 row: C column: 24
                Seq primer: CGTTGTAACGACGCGCCAGT
                Class: plasmid ends
                High quality sequence stop: 20.
FEATURES         Location/Qualifiers
                1..20
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUC1M0311C24"
                /clone_lib="Mouse 10kb plasmid UUC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                /note="vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptored DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pWD42 (gil4732114[gb|AF129072.1]), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptored mouse DNA was annealed to
                adaptored vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
BASE COUNT       2 a      4 c      6 g      8 t
ORIGIN
Query Match      100.0%; Score 7; DB 173; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tggtaga 7
    |||||
Db 8 TGTGGA 14

RESULT 3
LOCUS      AZ361153      21 bp      DNA      GSS      02-OCT-2000
DEFINITION  IM0104A17R Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUC1M0104A17 R, DNA sequence.
ACCESSION  AZ361153
VERSION    AZ361153.1 GI:10474853

```

VERSION KEYWORDS SOURCE	AZ450829.1 GSS. house mouse.	GI:10606020
ORGANISM	Mus musculus	
REFERENCE	1 (bases 1 to 21)	
AUTHORS	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.	
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0249 row: E column: 13 Seq primer: CACACAGGACACGCTATGACC Class: plasmid ends High quality sequence stop: 21.	
FEATURES	Location/Qualifiers	
source	1..21	
	/organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="UUCG1M0249E13" /clone_lib="Mouse 10kb plasmid UUCG1M library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-" /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g114732114 gb AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."	
BASE COUNT	10 a	8 a
ORIGIN	5 c	0 g
	6 t	5 t
Query Match	100.0%;	Score 7; DB 171; Length 21;
Best Local Similarity	100.0%;	Pred. No. 1.5e+05;
Matches	7; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 t g t t g g a 7	
Db	13 T G T T G G A 7	
RESULT	4	
LOCUS	AZ450829/c	21 bp DNA GSS 04-OCT-2000
DEFINITION	1M0249E13R Mouse 10kb plasmid UUCG1M library Mus musculus genomic clone UUCG1M0249E13 R, DNA sequence.	
ACCESSION	AZ450829	
Query Match	100.0%;	Score 7; DB 172; Length 21;
Best Local Similarity	100.0%;	Pred. No. 1.5e+05;
Matches	7; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 t g t t g g a 7	
Db	19 T G T T G G A 13	
RESULT	5	
LOCUS	AZ508369	21 bp DNA GSS 05-OCT-2000
DEFINITION	1M0350A24R Mouse 10kb plasmid UUCG1M library Mus musculus genomic clone UUCG1M0350A24 R, DNA sequence.	

```

AZ508369          AZ508369.1   GI:10689781
ACCESSION        VERSION
KEYWORDS         SOURCE
ORGANISM         house mouse; Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 21)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0350 row: A column: 24
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 21.
Location/Qualifiers
1..21
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/cclone="UUCG1M0350A24"
/clone.lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gbAF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
5 a      1 c      7 g      8 t
BASE COUNT      5 a      1 c      7 g      8 t
ORIGIN
Query Match      100.0%; Score 7; DB 174; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7
|||||||
Db 3 TGTTGA 9

RESULT 6
AI074857       25 bp mRNA EST 24-SEP-1998
LOCUS
DEFINITION     oyt27b11.s1 Soares_senescent_fibroblasts_Nbhsf Homo sapiens CDNA
AUTHORS
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 21)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0350 row: A column: 24
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 21.
Location/Qualifiers
1..21
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/cclone="UUCG1M0350A24"
/clone.lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gbAF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
5 a      1 c      7 g      8 t
BASE COUNT      5 a      1 c      7 g      8 t
ORIGIN
Query Match      100.0%; Score 7; DB 174; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7
|||||||
Db 3 TGTTGA 9

RESULT 6
AI074857       25 bp mRNA EST 24-SEP-1998
LOCUS
DEFINITION     oyt27b11.s1 Soares_senescent_fibroblasts_Nbhsf Homo sapiens CDNA
AUTHORS
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 21)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0350 row: A column: 24
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 21.
Location/Qualifiers
1..21
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/cclone="UUCG1M0350A24"
/clone.lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gbAF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
5 a      1 c      7 g      8 t
BASE COUNT      5 a      1 c      7 g      8 t
ORIGIN
Query Match      100.0%; Score 7; DB 174; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7
|||||||
Db 3 TGTTGA 9

RESULT 6
AI074857       25 bp mRNA EST 24-SEP-1998
LOCUS
DEFINITION     oyt27b11.s1 Soares_senescent_fibroblasts_Nbhsf Homo sapiens CDNA
AUTHORS
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 21)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0350 row: A column: 24
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 21.
Location/Qualifiers
1..21
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/cclone="UUCG1M0350A24"
/clone.lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gbAF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
5 a      1 c      7 g      8 t
BASE COUNT      5 a      1 c      7 g      8 t
ORIGIN
Query Match      100.0%; Score 7; DB 174; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7
|||||||
Db 3 TGTTGA 9

RESULT 6
AI074857       25 bp mRNA EST 24-SEP-1998
LOCUS
DEFINITION     oyt27b11.s1 Soares_senescent_fibroblasts_Nbhsf Homo sapiens CDNA
AUTHORS
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 21)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5
```

```

AZ508369          AZ508369.1   GI:10689781
ACCESSION        VERSION
KEYWORDS         SOURCE
ORGANISM         house mouse;
                 Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 21)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0350 row: A column: 24
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 21.
Location/Qualifiers
1..21
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/cclone="UUCG1M0350A24"
/clone.lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gbAF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
5 a      1 c      7 g      8 t
BASE COUNT      5 a      1 c      7 g      8 t
ORIGIN
Query Match      100.0%; Score 7; DB 174; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7
|||||||
Db 3 TGTTGA 9

RESULT 6
AI074857       25 bp mRNA EST 24-SEP-1998
LOCUS
DEFINITION     oyt27b11.s1 Soares_senescent_fibroblasts_Nbhsf Homo sapiens CDNA
AUTHORS
REFERENCE
AI074857       AI074857
LOCUS
DEFINITION     u42e07.x1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1885092 3' similar to TR:Q90658 Q90658 ZINC FINGER PROTEIN.
;, mRNA sequence.
AI158948
AI158948.1    GI:3692130
EST.
SOURCE
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 25)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

```

Wed Mar 28 14:03:19 2001

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Maria M/Mouse EST Project

WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:969416

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: custom primer used

High quality sequence stop: 1.

FEATURES  
source

1. .25  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1885092"  
/clone\_lib="Sugano mouse embryo mewa"  
/dev\_stage="embryo, 14 dpc"  
/lab\_host="DH10B"  
/note="vector: pME18S-FL3; Site\_1: DraIII (CACTGTGCTG); Site\_2: DraIII (CACCATGCG); 1st strand cDNA was primed with an oligo(dT) primer (ATGCGCCCTTTTTTTTTTTTTT); double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCACTGCG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGCTG, 3' site CACCATGCG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGCGAGCTCGAGCACA."

BASE COUNT 7 a 5 c 5 g 8 t

ORIGIN

Query Match 100.0%; Score 7; DB 16; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttga 7  
|||||  
Db 9 TGTGGA 15

RESULT 8  
AI453394/c  
LOCUS tJ37902.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2143730 3'  
DEFINITION similar to TR:Q39949 Q39949 HYDROXYPROLINE-RICH PROTEIN.; contains element MSK1 repetitive element.; mRNA sequence.

ACCESSION AI453394.1 GI:4281559

VERSION EST.

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 25)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Life Technologies catalog #: 11548-013  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
Insert Length: 1074 Std Error: 0.00  
Seq primer: 400p from Gibco

High quality sequence stop: 1.

FEATURES  
source

1. .25  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2143730"  
/clone\_lib="NCI\_CGAP\_Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.72 kb. Life Technologies catalog #:  
11548-013"

BASE COUNT 8 a 16 c 0 g 1 t

ORIGIN

Query Match 100.0%; Score 7; DB 20; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttga 7  
|||||  
Db 8 TGTGGA 2

RESULT 9

AZ352012/c

LOCUS

DEFINITION AZ352012 26 bp DNA GSS 29-SEP-2000  
clone UUGC1M0090M13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION AZ352012

VERSION AZ352012.1 GI:10431249

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 26)

REFERENCE

AUTHORS

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

JOURNAL

COMMENT

Contact: Robert B. Weiss

University of Utah

Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0090 row: M column: 13

Seq primer: CTTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 26.

Location/Qualifiers

1. 26

/organism="Mus musculus"

/strain="C57BL/6J"

```

/db_xref="taxon:10090"
/clone="UUGC1M0090M13"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      8 a      2 g      6 t
ORIGIN

Query Match      100.0%; Score 7; DB 171; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 t g t t g g a 7
    | | | | | | |
Db 19 T G T T G G A 13

RESULT 10
A2332711
LOCUS      29 bp      DNA      GSS      29-SEP-2000
DEFINITION clone UUGC1M0061N14 F, DNA sequence.
ACCESSION  A2332711
VERSION     A2332711.1 GI:10396624
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 29)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0061 row: N column: 14
            Seq primer: CTTGTAAACGACGCGCAGT
            Class: plasmid ends
            High quality sequence stop: 29.
            Location/Qualifiers
                1..29
                /organism="Mus musculus"

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0061N14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      5 a      4 c      9 g      11 t
ORIGIN

Query Match      100.0%; Score 7; DB 170; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 t g t t g g a 7
    | | | | | | |
Db 3 T G T T G G A 9

RESULT 11
A2419519/c
LOCUS      29 bp      DNA      GSS      03-OCT-2000
DEFINITION clone UUGC1M0196N03 F, DNA sequence.
ACCESSION  A2419519
VERSION     A2419519.1 GI:10543532
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 29)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0196 row: N column: 03
            Seq primer: CTTGTAAACGACGCGCAGT
            Class: plasmid ends
            High quality sequence stop: 29.
            Location/Qualifiers
                1..29
                /source

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0196N03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      9 a      8 c      5 g      7 t
ORIGIN

Query Match      100.0%; Score 7; DB 172; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttggg 7
    |||||
Db 17 TGTGGA 11

RESULT 12
AZ333172/c
LOCUS
DEFINITION
  LM0062B09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
  clone UUGC1M0062B09 F, DNA sequence.
ACCESSION
  AZ333172
VERSION
  AZ333172.1 GI:10397530
KEYWORDS
  GSS.
SOURCE
  house mouse.
  Mus musculus
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 31)
REFERENCE
  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
  Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
  M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
  and Wright,D., Weiss,R.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
  Unpublished (2000)
  Contact: Robert B. Weiss
  University of Utah Genome Center
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0062 row: B column: 09
  Seq primer: CGTTGTAACGACGCCAGT
  Class: plasmid ends
  High quality sequence stop: 31.
  Location/Qualifiers

BASE COUNT      9 a      8 c      5 g      7 t
ORIGIN

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0196N03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      9 a      8 c      5 g      7 t
ORIGIN

Query Match      100.0%; Score 7; DB 172; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttggg 7
    |||||
Db 17 TGTGGA 11

RESULT 12
AZ333172/c
LOCUS
DEFINITION
  LM0062B09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
  clone UUGC1M0062B09 F, DNA sequence.
ACCESSION
  AZ333172
VERSION
  AZ333172.1 GI:10397530
KEYWORDS
  GSS.
SOURCE
  house mouse.
  Mus musculus
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 31)
REFERENCE
  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
  Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
  M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
  and Wright,D., Weiss,R.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
  Unpublished (2000)
  Contact: Robert B. Weiss
  University of Utah Genome Center
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0062 row: B column: 09
  Seq primer: CGTTGTAACGACGCCAGT
  Class: plasmid ends
  High quality sequence stop: 31.
  Location/Qualifiers

BASE COUNT      11 a      12 c      4 g      4 t
ORIGIN

Query Match      100.0%; Score 7; DB 170; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttggg 7
    |||||
Db 14 TGTGGA 8

RESULT 13
AZ357647/c
LOCUS
DEFINITION
  AZ357647 Mouse 10kb plasmid UUGC1M library Mus musculus genomic
  clone UUGC1M0095D19 F, DNA sequence.
ACCESSION
  AZ357647
VERSION
  AZ357647.1 GI:10471347
KEYWORDS
  GSS.
SOURCE
  house mouse.
  Mus musculus
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 31)
REFERENCE
  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
  Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
  M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
  and Wright,D., Weiss,R.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
  Unpublished (2000)
  Contact: Robert B. Weiss
  University of Utah Genome Center
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0099 row: D column: 19
  Seq primer: CGTTGTAACGACGCCAGT
  Class: plasmid ends
  High quality sequence stop: 31.

```

FEATURES  
source

Location/Qualifiers  
1. .31  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCLM0099D19"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

FEATURES  
source

High quality sequence stop: 32.  
Location/Qualifiers  
1. .32  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCLM0040P07"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

10 a  
12 c  
5 g  
4 t

Query Match 100.0%; Score 7; DB 171; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7  
|||||  
Db 13 TGTGGGA 7

RESULT 14  
A3320254/c

LOCUS A3320254 32 bp DNA GSS 29-SEP-2000  
DEFINITION IM0040P07F Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0040P07 F, DNA sequence.

ACCESSION A3320254  
VERSION A3320254.1 GI:10371848

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 32)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0040 row: P column: 07  
Seq primer: CGTTGTAAACGCGCCAGT  
Class: plasmid ends

BASE COUNT  
ORIGIN

9 a  
11 c  
3 g  
9 t

Query Match 100.0%; Score 7; DB 170; Length 32;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7  
|||||  
Db 22 TGTGGGA 16

## RESULT .15

A3328463  
LOCUS A3328463

DEFINITION AZ328463 32 bp DNA GSS 29-SEP-2000  
clone UUGCLM0052D18 F, DNA sequence.

ACCESSION AZ328463  
VERSION AZ328463.1 GI:10388217

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 32)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0052 row: D column: 18  
Seq primer: CGTTGTAAACGCGCCAGT

Class: plasmid ends  
High quality sequence stop: 32.

FEATURES  
Location/Qualifiers  
1..32  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0052D18"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWBx2 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 7 a 3 c 12 g 10 t  
ORIGIN

Query Match 100.0%; Score 7; DB 170; Length 32;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggg 7  
      |||||  
Db 7 TGTGGA 13

Search completed: March 27, 2001, 07:38:21  
Job time: 4584 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:17 ; Search time 2517.78 Seconds  
(without alignments)  
61.230 Million cell updates/sec

Title: US-09-380-826A-2  
Perfect score: 22  
Sequence: 1 ttgttgatcacaaagtattgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estov1.\*  
68: em\_estov2.\*  
69: em\_estpl1.\*  
70: em\_estpl2.\*  
71: em\_estpl3.\*  
72: em\_estpl4.\*  
73: em\_estpl5.\*  
74: em\_estro1.\*  
75: em\_estro2.\*  
76: em\_estro3.\*  
77: em\_estro4.\*  
78: em\_estro5.\*  
79: em\_estro6.\*  
80: em\_estro7.\*  
81: em\_estro8.\*  
82: em\_estro9.\*  
83: em\_estro10.\*  
84: em\_estro11.\*  
85: em\_estro12.\*  
86: em\_estro13.\*  
87: gb\_est41.\*  
88: gb\_est42.\*  
89: gb\_est43.\*  
90: gb\_est44.\*  
91: gb\_est45.\*  
92: gb\_est46.\*  
93: gb\_est47.\*  
94: gb\_est48.\*  
95: gb\_est49.\*  
96: gb\_est50.\*  
97: gb\_est51.\*  
98: gb\_est52.\*  
99: gb\_est53.\*  
100: gb\_est54.\*  
101: gb\_est55.\*  
102: gb\_est56.\*  
103: gb\_est57.\*  
104: gb\_est58.\*  
105: gb\_est59.\*  
106: gb\_est60.\*  
107: gb\_est61.\*  
108: gb\_est62.\*  
109: gb\_est63.\*  
110: gb\_est64.\*  
111: gb\_est65.\*  
112: em\_esthum21.\*  
113: em\_esthum22.\*  
114: em\_esthum23.\*  
115: em\_estom1.\*  
116: em\_estom2.\*

117: em\_estpl6:\*  
118: em\_estpl7:\*  
119: em\_estpl8:\*  
120: em\_estpl9:\*  
121: em\_estpl10:\*  
122: em\_estpl11:\*  
123: em\_estpl12:\*  
124: em\_estpl13:\*  
125: em\_estpl14:\*  
126: em\_estpl15:\*  
127: em\_estpl16:\*  
128: em\_estpl17:\*  
129: em\_estpl18:\*  
130: em\_estpl19:\*  
131: em\_estpl20:\*  
132: em\_estpl21:\*  
133: em\_estpl22:\*  
134: em\_estpl23:\*  
135: em\_estpl24:\*  
136: em\_estpl25:\*  
137: em\_estpl26:\*  
138: em\_estpl27:\*  
139: em\_estpl28:\*  
140: em\_estpl29:\*  
141: em\_estpl30:\*  
142: em\_estpl31:\*  
143: em\_estpl32:\*  
144: em\_estpl33:\*  
145: em\_estpl34:\*  
146: em\_estpl35:\*  
147: em\_estpl36:\*  
148: em\_estpl37:\*  
149: em\_estpl38:\*  
150: em\_estpl39:\*  
151: em\_estpl40:\*  
152: em\_estpl41:\*  
153: em\_estpl42:\*  
154: em\_estpl43:\*  
155: em\_estpl44:\*  
156: em\_estpl45:\*  
157: em\_estpl46:\*  
158: em\_estpl47:\*  
159: em\_estpl48:\*  
160: em\_estpl49:\*  
161: em\_estpl50:\*  
162: em\_estpl51:\*  
163: em\_estpl52:\*  
164: em\_estpl53:\*  
165: em\_estpl54:\*  
166: em\_estpl55:\*  
167: em\_estpl56:\*  
168: em\_estpl57:\*  
169: em\_estpl58:\*  
170: em\_estpl59:\*  
171: em\_estpl60:\*  
172: em\_estpl61:\*  
173: em\_estpl62:\*  
174: em\_estpl63:\*  
175: em\_estpl64:\*  
176: em\_estpl65:\*  
177: em\_estpl66:\*  
178: em\_estpl67:\*  
179: em\_estpl68:\*  
180: em\_estpl69:\*  
181: em\_estpl70:\*  
182: em\_estpl71:\*  
183: em\_estpl72:\*  
184: em\_estpl73:\*  
185: em\_estpl74:\*  
186: em\_estpl75:\*  
187: em\_estpl76:\*  
188: em\_estpl77:\*  
189: em\_estpl78:\*

190: gb\_gss25:\*  
191: gb\_gss26:\*  
192: gb\_gss27:\*  
193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.8	85.5	361	37	AV532040	AV532040 AV532040
C 2	18.8	85.5	394	147	Z26559	Z26559 ATTS1651 ve
C 3	18.8	85.5	490	14	AB038725	AB038725 AB038725
C 4	18.8	85.5	527	37	AV520789	AV520789 AV520789
C 5	18.8	85.5	530	27	AI995637	AI995637 701676626
C 6	18.4	83.6	608	138	BE977618	BE977618 bs66108.y
C 7	18.4	83.6	934	191	CNS02BWA	AL190387 Tetraodon
C 8	17.8	80.9	600	138	BF006444	BF006444 EST434942
C 9	17.8	80.9	633	138	BF006254	BF006254 EST434942
C 10	17.4	79.1	286	126	BB288727	BB288727 BB288727
C 11	17.4	79.1	407	110	BE581720	BE581720 kg51c07.y
C 12	17.4	79.1	529	158	AQ497779	AQ497779 HS_5066.B
C 13	17.4	79.1	674	94	AW695022	AW695022 NF082804S
C 14	17.4	79.1	762	106	BE283043	BE283043 601101323
C 15	17.4	79.1	837	110	BE642809	BE642809 Cr12_7.B0
C 16	17.4	79.1	1101	190	CNS00D26	AL077427 Drosophila
C 17	17.2	78.2	151	134	BE068225	BE068225 MR4-BF036
C 18	17.2	78.2	425	15	A1049911	A1049911 an30h03.x
C 19	17.2	78.2	431	159	AQ595814	AQ595814 HS_2132.B
C 20	17.2	78.2	432	150	AQ221715	AQ221715 HS_2010.A
C 21	17.2	78.2	437	146	W43212	W43212 22591 Lamd
C 22	17.2	78.2	439	162	AQ796989	AQ796989 nbdp0071H
C 23	17.2	78.2	500	24	AI733664	AI733664 an30h03.x
C 24	17.2	78.2	524	174	B62582	B62582 T22F18TR TA
C 25	17.2	78.2	531	150	AQ223477	AQ223477 HS_2003.B
C 26	17.2	78.2	536	39	AQ034253	AW034253 EST377824
C 27	17.2	78.2	546	24	AI779714	AI779714 EST260593
C 28	17.2	78.2	579	174	B67515	B67515 T22M10TR TA
C 29	17.2	78.2	602	24	AI777095	AI777095 EST358060
C 30	17.2	78.2	614	105	BE187570	BE187570 EST336131
C 31	17.2	78.2	619	97	AW963686	AW963686 EST375759
C 32	17.2	78.2	680	174	B57784	B57784 CIT-HSP-201
C 33	17.2	78.2	698	168	A2193715	A2193715 SP_1023.B
C 34	17.2	78.2	904	191	CNS029DN	AL187124 Tetraodon
C 35	17.2	78.2	949	192	CNS04HNP	AL291166 Tetraodon
C 36	17.2	78.2	373	140	C70416	C70416 C70416 Yuj1
C 37	16.8	76.4	238	34	AV313332	AV313332 AV313332
C 38	16.8	76.4	260	104	BE118999	BE118999 UI-R-CA0-
C 39	16.8	76.4	296	145	T20362	T20362 6c01908-t7
C 40	16.8	76.4	396	97	AW943911	AW943911 LD47517.3
C 41	16.8	76.4	397	91	AW487226	AW487226 81727 MAR
C 42	16.8	76.4	432	157	AQ449764	AQ449764 500004C01
C 43	16.8	76.4	446	24	AI761541	AI761541 w161f07.x
C 44	16.8	76.4	450	138	BF002933	BF002933 7g51c09.x
C 45	16.8	76.4	464	156	AQ332265	AQ332265 HS_5007_A

ALIGNMENTS

RESULT 1  
AV532040/c  
LOCUS AV532040 Arabidopsis thaliana flower buds Columbia Arabidopsis.  
DEFINITION thaliana cDNA clone FB034f09F 3', mRNA sequence.  
ACCESSION AV532040  
VERSION AV532040.1 GI:8692323  
KEYWORDS EST.  
SOURCE thale cress.

AV532040 361 bp mRNA EST 01-SEP-2000

ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 361)  
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

JOURNAL DNA Res. 7, 175-180 (2000)

MEDLINE 20363093

COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source

1..361 /organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="FB034f09f"

/clone\_lib="Arabidopsis thaliana flower buds Columbia"

/tissue\_type="flower buds"

/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:

XhoI"

BASE COUNT 113 a 73 c 61 g 114 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 37; Length 361;

Best Local Similarity 90.9%; Pred. No. 34;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22

||||| |||||||||

Db 276 TGTGGTTCACAGATTTCACA 255

RESULT 2

226559/c

LOCUS

226559 394 bp mRNA EST 31-MAY-1995

DEFINITION ARTS1651 Versailles-VC Arabidopsis thaliana cDNA clone VCVDH08 3'

similar to Ribonuclease (RNS2), mRNA sequence.

ACCESSION 226559

VERSION 226559.1 GI:404227

KEYWORDS EST.

SOURCE

thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 394)

CNRS.

The Arabidopsis thaliana transcribed genome: the GDR cDNA program

Unpublished (1996)

JOURNAL

COMMENT Contact: Desprez T., Anselem J., Chiapello H., Rouze P., Caboche

M., Hofte H.

INRA Versailles

Laboratoire de Biologie Cellulaire

Route de Saint-Cyr, 78026 Versailles Cedex, France

Email: thierry@versailles.inra.fr.

Location/Qualifiers

1..394

/organism="Arabidopsis thaliana"

/strain="ecotype Columbia"

/db\_xref="taxon:3702"

/clone="VCVDH08"

/clone\_lib="Versailles-VC"

/tissue\_type="whole seedlings"

/dev\_stage="in vitro-grown etiolated seedlings, 5 days old"

/note="Vector: pBluescript"

BASE COUNT 117 a 80 c 63 g 133 t

1 others

ORIGIN

thale cress.

ORGANISM Arabidopsis thaliana

ORIGIN

Query Match 85.5%; Score 18.8; DB 147; Length 394;

Best Local Similarity 90.9%; Pred. No. 35;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacaaagatttgata 22

||||| |||||||||

Db 321 TGTGGTTCACAGATTTCACA 300

RESULT 3

AB038725/c

LOCUS

AB038725 490 bp mRNA EST 29-SEP-2000

DEFINITION AB038725 Arabidopsis thaliana Above-ground organ from two to

six-week old plants Columbia Arabidopsis thaliana cDNA clone

AP230903\_f 3', mRNA sequence.

ACCESSION AB038725

VERSION AB038725.1 GI:7212552

KEYWORDS EST.

SOURCE

thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 490)

AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation

of 12,028 non-redundant expressed sequence tags from normalized and

size-selected cDNA libraries

JOURNAL DNA Res. 7, 175-180 (2000)

MEDLINE 20363093

COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers

1..490

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="AP230903\_f"

/clone\_lib="Arabidopsis thaliana Above-ground organ from

two to six-week old plants Columbia"

/tissue\_type="Above-ground organ from two to six-week old

plants"

/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:

XhoI"

BASE COUNT 147 a 99 c 87 g 157 t

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0;

Qy 1 tgttgatcacaaagatttgata 22

||||| |||||||||

Db 303 TGTGGTTCACAGATTTCACA 282

RESULT 4

AV520789/c

LOCUS

AV520789 527 bp mRNA EST 01-SEP-2000

DEFINITION AV520789 Arabidopsis thaliana aboveground organs two to six-week

old Arabidopsis thaliana cDNA clone AP232906F 3', mRNA sequence.

ACCESSION AV520789

VERSION AV520789

KEYWORDS EST.

SOURCE

thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

## REFERENCE

1 (bases 1 to 527)

ASAMIZU, E., NAKAMURA, Y., SATO, S. and TABATA, S.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

## JOURNAL

## MEDLINE

## COMMENT

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizukazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

## FEATURES

source

1..527

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone\_lib="AP232g06f"

/clone\_lib="Arabidopsi\$ thaliana aboveground organs two to six-week old"

/tissue\_type="aboveground organs"

/dev\_stage="two to six-week old"

/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

157 a 111 c 101 g 158 t

## BASE COUNT

ORIGIN

Query Match 85.5%; Score 18.8; DB 37; Length 527;  
Best Local Similarity 90.9%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacagatttgata 22

||||| |||||||||

Db 247 TGTGGTTCACAGATTGACA 226

RESULT 5  
AI995637/c 530 bp mRNA EST 08-SEP-1999  
LOCUS 701676626 A. thaliana, Columbia Col-0, inflorescence-1 Arabidopsis  
DEFINITION thaliana cDNA clone 701676626, mRNA sequence.

ACCESSION AI995637

VERSION AI995637.1 GI:5842542

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

## REFERENCE

## AUTHORS

Chen, J., Momiya, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guebler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, B., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobriaga, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.

Arabidopsis thaliana Gene Expression MicroArray

Unpublished (1999)

Contact: David Smoller, Ph.D.

Genome Systems, Inc., a wholly owned subsidiary of Incyte

Pharmaceuticals, Inc.

4633 World Parkway Circle, St. Louis, MO 63134, USA

Tel: 877-577-2733

Fax: 314-427-3324

Email: [service@genomesystems.com](mailto:service@genomesystems.com).

Location/Qualifiers

1..530

/organism="Arabidopsis thaliana"

/cultivar="Columbia Col-0"

## FEATURES

source

/db\_xref="taxon:3702"

/clone\_lib="A. thaliana, Columbia Col-0, inflorescence-1"

/tissue\_type="inflorescence"

/dev\_stage="4 - 7 weeks"

/note="Vector: pSPORT; Site\_1: NotI; Site\_2: SalI; cDNA

library was derived from untreated inflorescence tissue

from Arabidopsis thaliana, Columbia Col-0, at 4 - 7

weeks. Plants were grown in 1:1:1 peat

moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C

under constant light, and watered with fertilizer. cDNA

synthesis was initiated using a NotI-oligo(dT) primer.

Double-stranded cDNA was blunt-ended, ligated to SalI adaptors

, digested with NotI, size-selected, and cloned into the

NotI and SalI sites of the pSPORT vector."

163 a 108 c 100 g 158 t 1 others

BASE COUNT 100 g 158 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 27; Length 530;  
Best Local Similarity 90.9%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgttgatcacagatttgata 22

||||| |||||||||

Db 254 TGTGGTTCACAGATTGACA 233

## RESULT

6

BE977618/c

LOCUS BE977618

DEFINITION BE977618

ACCESSION BE977618

VERSION BE977618.1

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 608)

ANDREWS, J., BOUFFARD, G. and OLIVER, B.

Drosophila melanogaster testis expressed sequence tags

Unpublished (1999)

Contact: Brian Oliver

Laboratory of Cellular and Developmental Biology

NIDDK, National Institutes of Health

6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA

Fax: (301) 496 5239

Email: [oliver@helix.nih.gov](mailto:oliver@helix.nih.gov),

<http://www.nidk.nih.gov/intram/people/bolliver.htm>

Tissue isolation and library construction performed at the National

Institute of Diabetes and Digestive and Kidney Diseases, NIH (see

<http://www.nidk.nih.gov/intram/people/bolliver.htm>). DNA sequencing

and analyses performed by National Institutes of Health Intramural

Sequencing Center (NISC; see <http://www.nisc.nih.gov>).

Plate: 66 row: h column: 08

Seq primer: M13KPL reverse primer (ABI).

Location/Qualifiers

1..608

/organism="Drosophila melanogaster"

/strain="y[\*] w[67c1]/Y"

/db\_xref="taxon:722"

/clone\_lib="bs66h08"

/sex="male"

/dev\_stage="1-5 day adult"

/lab\_host="SOLR (Stratagene)"

/note="Organ: testis; Vector: pBluescript SK (Stratagene);

Site\_1: EcoRI; Site\_2: Xho I; Testes dissected from 1-5

day adult y[\*] w[67c1]/Y males raised at 25°C. RNA

isolated using Trizol (Life Technologies) and a single

round of Poly(A)+ selection using Oligotex (Qiagen). cDNA library constructed using Stratagene ZAP-cDNA synthesis kit. Oligo dT-primed, size fractionated -1-6 kb, and directionally cloned at EcoRI and XhoI in Uni-ZAP XR. Following a single round of amplification pBluescript SK phagemids were mass excised. A distribution channel for clones is being sought, but not currently available. Requests for clones cannot be honored."

BASE COUNT 173 a 146 c 161 g 128 t  
ORIGIN

Query Match 83.6%; Score 18.4; DB 138; Length 608;  
Best Local Similarity 95.0%; Pred. No. 59;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagtattgata 22  
||||| |||||||||  
Db 74 TTGGATCCCAAGATTGATA 55

RESULT 7  
LOCUS CNS02BWA/c 934 bp DNA GSS 12-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 254C22 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL190387  
VERSION AL190387.1 GI:7828491  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis

REFERENCE 1 (bases 1 to 934)  
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.  
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 934)  
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 934)  
AUTHORS Genoscope.

TITLE Direct Submission  
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
source  
1. 934  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="254C22"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG254BB11SP1-end : PUC-Ori"

BASE COUNT 282 a 175 c 207 g 259 t 11 others  
ORIGIN

Query Match 83.6%; Score 18.4; DB 191; Length 934;  
Best Local Similarity 95.0%; Pred. No. 62;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 3 ttggatcacaaagtattgata 22  
||||| |||||||||  
Db 840 TTGGACACACAGATTGATA 821

RESULT 8  
LOCUS BF006444/c 600 bp mRNA EST 06-OCT-2000  
DEFINITION EST434942 DSLC Medicago truncatula cDNA clone pDSLCL-41J15, mRNA sequence.

ACCESSION BF006444  
VERSION BF006444.1 GI:10706719  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 600)  
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Gonzales,M.B. and Ellis,L.  
TITLE ESTs from Medicago truncatula leaves and cotyledons

JOURNAL Unpublished (2000)  
COMMENT Contact: Deborah A. Samac  
Department of Plant Pathology  
University of Minnesota  
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA  
Tel: 612 625 1243  
Fax: 651 649 5058

Email: debbye@puccini.crl.umn.edu  
University of Minnesota name: M275482e TIGR sequence name: MTLBA56TK More information is available at: <http://chrysis.tamu.edu/medicago>  
Seq primer: SKmod (CTA GAA CTA gta gat CC).

FEATURES  
source  
1. 600  
/organism="Medicago truncatula"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="pDSLCL-41J15"  
/clone\_lib="DSLC"  
/tissue\_type="leaves and cotyledons"  
/dev\_stage="mixture of cotyledons from five days old plants and leaves obtained from two weeks old plants"  
/lab\_host="E. coli strain SOLR"  
/note="Vector: pBluescript SK +/-; Site1: EcoRI; Site2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of cotyledons of five days old plants and leaves of two weeks old plants. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 127 a 120 c 132 g 221 t  
ORIGIN

Query Match 80.9%; Score 17.8; DB 138; Length 600;  
Best Local Similarity 90.5%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tggatcacaaagtattgata 21  
||||| |||||||||  
Db 90 TCTTGAATCACAGATTAGAT 70

RESULT 9  
LOCUS BF006254/c 633 bp mRNA EST 06-OCT-2000  
DEFINITION EST434752 DSLC Medicago truncatula cDNA clone pDSLCL-40G21, mRNA

```

sequence.
ACCESSION   BF006254
VERSION     BF006254.1  GI:10706529
KEYWORDS    EST.
SOURCE      barrel medic.
ORGANISM    Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1 (bases 1 to 633)
AUTHORS     Fedorova, M., Pierson, B.L., Samac, D.A., Gantt, J.S., Vance, C.P.,
            Gonzales, M.B. and Ellis, L.
            ESTs from Medicago truncatula leaves and cotyledons
            Unpublished (2000)
            Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 612 649 5058
            Email: debbys@pcc.umn.edu
            University of Minnesota name: M275292e TIGR sequence name:
            MTLA047PK More information is available at:
            http://chrystle.tamu.edu/medicago
            Seq primer: Sknod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1..633
                     /organism="Medicago truncatula"
                     /cultivar="genotype Al7"
                     /db_xref="taxon:3880"
                     /clone_lib="pSLC-40621"
                     /clone="pSLC-40621"
                     /tissue_type="leaves and cotyledons"
                     /dev_stage="mixture of cotyledons from five days old
                     plants and leaves obtained from two weeks old plants"
                     /lab_host="E. coli strain SOLR"
                     /note="vector: phluescript SK +/-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from the
                     mixture of cotyledons of five days old plants and leaves
                     of two weeks old plants. The cDNA was directionally
                     ligated into the Uni-ZAP XR vector from Stratagene and
                     packaged using Gigapack III Gold packaging extracts.
                     Plasmids containing cDNA inserts were excised from the
                     recombinant lambda-ZAP phage using Ex-Assist helper phage
                     and propagated in SOLR cells."
BASE COUNT        148 a 132 c 124 g 229 t
ORIGIN
1 ttttgatcacaaagtattgat 21
Db 188 TCTTGATCACAGATTAGAT 168

Query Match      80.98; Score 17.8; DB 138; Length 633;
Best Local Similarity 90.5%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttttgatcacaaagtattgat 21
Db 188 TCTTGATCACAGATTAGAT 168

RESULT 10
BB288727/c
LOCUS       BB288727       286 bp      mRNA      EST      09-JUL-2000
DEFINITION   BB288727 RIKEN full-length enriched, 2 cells egg Mus musculus cDNA
            clone B020032C12 3', mRNA sequence.
ACCESSION   BB288727
VERSION     BB288727.1  GI:8989176
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 286)
REFERENCE   1 (bases 1 to 286)
AUTHORS     Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
            P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
            Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
            Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
            Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
            Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
            Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata
            Y., Shigenoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
            Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya
            T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
            Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
            M., Muramatsu, M. and Hayashizaki, Y.
            RIKEN Mouse ESTs (Konno, H., et al.)
            Unpublished (2000)
            Contact: Toshinide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL: http://genome.rtc.riken.go.jp/
            Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
            N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
            Thermotabilization and thermoinactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
            Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
            Y. and Hayashizaki, Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci, P. and Hayashizaki, Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..286
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="B020032C12"
                     /clone_lib="RIKEN full-length enriched, 2 cells egg"
                     /tissue_type="egg"
                     /dev_stage="2 cells"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTT 3']. cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGATTCCTCGATTAAATAATTAATCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT        117 a 66 c 20 g 83 t
ORIGIN
84 gttggatcacaaagtattga 20
117 gttggatcacaaagtattga 151
169 gttggatcacaaagtattga 151

Query Match      79.1%; Score 17.4; DB 126; Length 286;
Best Local Similarity 94.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagtattga 20
117 gttggatcacaaagtattga 151
169 gttggatcacaaagtattga 151

```

```

RESULT 11
BE581720/c 407 bp mRNA EST 16-AUG-2000
LOCUS Kq51c07.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to
DEFINITION WP.T19810.2 CE16413 ;, mRNA sequence.
ACCESSION BE581720
VERSION BE581720.1 GI:9832662
KEYWORDS EST.
SOURCE Strongyloides stercoralis.
ORGANISM Strongyloides stercoralis.

REFERENCE
AUTHORS Eukaryota: Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimidae; Strongyloidea; Strongyloides.
1 (bases 1 to 407)
McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas
M., McCann,R., Waterston,R. and Wilson,R. 1999
The Washington Univ. Nematode EST Project, 1999
Contact: McCarter JP
Unpublished (1999)
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of
NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
High quality sequence stop: 384.
FEATURES
source
1. .407
/organism="Strongyloides stercoralis"
/strain="Rhabditiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/clone_lib="TBN95TM-SSR"
/lab_host="XL-1 Blue MRF" (Stratagene)"
/notes="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
rhabditiform larvae which had been isolated from gerbils
experimentally infected with larvae originally isolated
from experimentally infected dogs. cDNA was constructed
and, using adaptors, was cloned unidirectionally into the
vector from the EcoRI site to the XhoI site. The library
has an unamplified titer of 1 x 10E5 pfu/ml and an
amplified, undiluted titer of 9 x 10E11 pfu/ml. The
average insert size of the unamplified library is 675 bp
(range, 100-1700)."
BASE COUNT 152 a 63 c 74 g 118 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 110; Length 407;
Best Local Similarity 94.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ttgatcacaaagatttgata 22
|||||
Db 334 TTGATCACAGATTGTGATA 316

RESULT 12
A0497779
LOCUS HS_5066_B1_E05_37A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone J7A-642 Col-9 Row-J, DNA sequence.
ACCESSION A0497779
VERSION A0497779.1 GI:4697902
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 529)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 642 row: J column: 9
Seq primer: T7
Class: BAC ends
High quality sequence stop: 529.
FEATURES
source
1. .529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-642 Col-9 Row-J"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 145 a 99 c 109 g 165 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 158; Length 529;
Best Local Similarity 94.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagatttgat 21
|||||
Db 403 TTGGTCACAGATTGTGAT 421

RESULT 13
AW695022/c
LOCUS AW695022 674 bp mRNA EST 15-JUN-2000
DEFINITION NF082E04ST1F1034 Developing stem Medicago truncatula cDNA clone
NF082E04ST 5', mRNA sequence.
ACCESSION AW695022
VERSION AW695022
KEYWORDS EST.
SOURCE AW695022.1 GI:7569784
ORGANISM barrel medic.
Medicago truncatula
Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 674)
He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
,R.A.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula stem library
Unpublished (2000)
Contact: Dixon RA

```

Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7302  
Fax: 580 221 7380  
Email: radixon@noble.org

Insert Length: 674 Std Error: 0.00  
Plate: 082 row: E column: 04  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers

# FEATURES

source

1. .674  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF082E045T"  
/tissue\_type="stem"  
/dev\_stage="pooled developmental"  
/note="Vector: Lambda Zap; Contains a mixture of  
internodal stem segments"

BASE COUNT 189 a 149 c 133 g 203 t

# ORIGIN

Query Match 79.1%; Score 17.4; DB 94; Length 674;  
Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttggatcacagaattgat 21  
|||||

Db 256 TTGGATCACTAGATTGAT 238

# RESULT 14

BE283043

LOCUS

DEFINITION BE283043 762 bp mRNA EST 13-JUL-2000  
601101323F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3493829 5',  
mRNA sequence.

ACCESSION BE283043

VERSION BE283043.1

KEYWORDS GI:9159023

SOURCE EST.

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 762)

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LIAM8541 row: n column: 06

High quality sequence stop: 661.

Location/Qualifiers

1. .762

/organism="Mus musculus"

/strain="C57BL/6J (f1eral)"

/db\_xref="taxon:10090"

/clone="IMAGE:3493829"

/tissue\_type="spontaneous tumor, metastatic to mammary."

/lab\_host="DH10B"

Stem cell origin.

/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: SalI;

Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

BASE COUNT 193 a 172 c 207 g 190 t

# ORIGIN

Query Match 79.1%; Score 17.4; DB 106; Length 762;  
Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaatttg 19  
|||||

Db 692 TGTGGATCACAGACTTG 710

# RESULT 15

BE642809

LOCUS

DEFINITION BE642809 837 bp mRNA EST 01-SEP-2000  
Crl2\_7\_B04\_SP6 Ceratopteris Spore Library Ceratopteris richardii  
cDNA clone Crl2\_7\_B04 5', mRNA sequence.

ACCESSION BE642809

VERSION BE642809.1

KEYWORDS GI:9960488

SOURCE EST.

ORGANISM

Ceratopteris richardii.

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Filicophyta;

Filicopsida; Filicales; Pteridaceae; Ceratopteris.

1 (bases 1 to 837)

Chatterjee, A., San Miguel, P., Stout, S.C., Banks, J. and Roux, S.J.

Expressed sequence tags of cDNA clones from a C. richardii library

Unpublished (2000)

Contact: Roux SJ

Section of Molecular Cell and Developmental Biology

University of Texas

Biology Building, Room 16, Austin, TX 78712, USA

Tel: 512 471 4238

Fax: 512 232 3402

Email: sroux@uts.cc.utexas.edu

Plate: Crl2\_7 row: B column: 04

Seq primer: SP6.

Location/Qualifiers

1. .837

/organism="Ceratopteris richardii"

/cultivar="Brogn"

/db\_xref="taxon:49495"

/clone="Crl2\_7\_B04"

/tissue\_type="Ceratopteris Spore Library"

/cell\_type="Spore"

/dev\_stage="20 hours after germination initiation"

/note="Vector: pCMVSPORT6; EST sequence from cDNA library.

cDNA library constructed from mRNA isolated from C.

richardii spores that had developed for 20 hours after

their germination had been initiated by white light."

BASE COUNT 233 a 158 c 214 g 232 t

ORIGIN

Query Match 79.1%; Score 17.4; DB 110; Length 837;

Best Local Similarity 94.7%; Pred. No. 1.9e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaatttg 19  
|||||

Db 332 TGTAGGATCACAGATTG 350

Search completed: March 27, 2001, 07:38:20  
Job time: 4583 sec



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 06:54:38 ; Search time 207.51 Seconds  
(without alignments)  
2673.865 Million cell updates/sec

Title: us-09-380-826A-1

Perfect score: 1477

Sequence: 1 gatcatggctggaactaac.....ccgtaaatcattctctgcag 1477

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

```
1: /cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1477	100.0	1477	19	V58896
2	814.6	55.2	1535	20	X83569
3	813.4	55.1	1512	20	X82000
4	811.6	54.9	1529	20	X83564
5	810.2	54.9	1508	20	Z31557
6	810.2	54.9	1508	21	Z45352
7	810	54.8	1529	20	X83567
8	808.4	54.7	1529	20	X83565
9	808.4	54.7	1529	20	X83566
10	807.8	54.7	1513	20	X81999
11	805.4	54.5	1521	16	Q92879
12	805.4	54.5	1526	16	T01866

13	805	54.5	1528	20	X83570	16S rDNA gene frag
14	805	54.5	1535	20	X83568	16S rDNA gene frag
15	803.8	54.4	1536	17	T18765	Pseudomonas testos
16	802.8	54.4	1477	20	Z10000	Eubacteria sp. 16S
17	801.8	54.3	1512	20	X81996	B. cereus almost c
18	801.6	54.3	1555	17	T29142	rRNA gene (Stp.aur
19	801.6	54.3	1555	19	V24294	Staphylococcus aur
20	801.4	54.3	1450	21	Z45351	Nucleotide sequenc
21	800.4	54.2	1532	13	Q26729	16S rRNA gene from
22	796.4	53.9	1506	20	Z31560	Bacillus species p
23	795.6	53.9	1532	15	Q64008	16S rRNA gene. Bo
24	794.8	53.8	1446	17	T10955	Root Stimulating B
25	794.8	53.8	1535	19	V24295	Burkholderia cepac
26	793.6	53.7	1535	20	X84802	Nitrosopira 16S rDN
27	791.8	53.6	1477	20	X26285	16S rDNA sequence
28	789.2	53.4	1452	20	X82004	Curtobacterium sp.
29	789	53.4	1535	20	X83571	16S rDNA gene frag
30	786.8	53.3	1477	17	T43670	Segmented filament
31	785.8	53.2	1455	20	X77384	B. subtilis AS2 st
32	784.4	53.1	1566	17	T18645	16S ribosomal RNA
33	781	52.9	1540	10	N91514	Escherichia coli 1
34	780.2	52.8	1542	17	T18759	E. coli 16S riboso
35	780	52.8	1542	17	T29140	rRNA gene (rrsE) f
36	780	52.8	5098	20	X24984	E. coli MG1655 rrr
37	780	52.8	5341	20	X24986	E. coli MG1655 rrr
38	779.8	52.8	1474	17	T18766	Pseudomonas cepaci
39	779.4	52.8	1548	20	X84804	Nitrosopira 16S rDN
40	778.4	52.7	1556	20	X81995	Sphingomonas 16S r
41	778.4	52.7	5097	20	X24983	E. coli MG1655 rrr
42	777.2	52.6	5105	20	X24989	E. coli MG1655 rrr
43	776.8	52.6	5014	20	X24987	E. coli MG1655 rrr
44	775.6	52.5	1513	20	X82003	B. megaterium ssp.
45	775.4	52.5	1450	20	X82005	Sphingomonas sp. B

#### ALIGNMENTS

RESULT 1

ID V58896 standard; DNA: 1477 BP.

XX V58896;

AC 20-JAN-1999 (first entry)

DT L. fainei nucleotide sequence.

DE Infection; pathogenic Leptospira; protective immunity; therapy;

KW diagnosis; ss.

OS Leptospira fainei.

PN WO9840099-A1.

PD 17-SEP-1998.

PF 06-MAR-1998; 98WO-AU00145.

PR 07-MAR-1997; 97AU-0005494.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

PA (PIGR-) PIG RES & DEV CORP.

XX Chappel RJ;

XX WPI; 1998-520791/44.

XX New isolated pathogenic Leptospira bacterium - useful for, e.g

PT developing products for conferring protective immunity, and for

XX prophylactic or therapeutic treatment

PS Claim 15; Page 69-70; 94pp; English.



















```

Db 600 gggctcaacctgggaactgattgaaactgacgagctagaatgcagtagagggaggtgg 659
QY 630 aattccaggtgtagcgggtgaaatgcgtagatatctgagagaacacacagtgccgaagcgga 689
Db 660 aattccggtgtagcgggtgaaatgcgtagatcgagagaacacacagtgccgaagcggtg 719
QY 690 cttgctggtctaaactgacgtgagcgacgaagacgctgggtgtagtaaaacggtattagata 749
Db 720 cctctcggtactgacattgacgtgagcgtggaagcgtgggagcaaacagattagata 779
QY 750 ccccggttaataccacccctaaactgtgtctaccagttttaggggtttta--acctcag 807
Db 780 cctggtagtcacacccgcgaacagtgctctactagctcgtgggaatcttagtattcttgg 839
QY 808 taacgaacctaacgattaaatagaccccgctgggactatgctcgaagagtgaaactca 867
Db 840 tgacgaagttaacgcgataagtagaccgctgggagtagcgcgcaaggttaaaactca 899
QY 868 aaggaattgacgggggtccgcacaagcgggtggagcatgtggttttaattcgatgatacccc 927
Db 900 aatgaattgacggggggccgcacaagcgggtggagcatgtggttttaattcgatgatacccc 959
QY 928 aaaaacctcaactgggcttgacatgattctga-atcatgtagagatatatgagccttcgg 986
Db 960 aagaaccttaccagccttgacatccttggaactttctagagatagattggtgccttcgg 1019
QY 987 --gcagattcacagtgctgattggttctgctcaactcgtgtcgtagatgttgggttaa 1044
Db 1020 gagccaagtgcacagtgctgattggttctgctcaactcgtgtcgtagatgttgggttaa 1079
QY 1045 gtccgcgaacgagcgcaacctctatctgttgc---taccttaagttggcactgta 1101
Db 1080 gtcccgtaacgagcgcaaccttgcctctagtgtccgaacacataatgggtgggaaccttag 1139
QY 1102 cgaacctccggtgacaaacccgggaagcggggatgacgtcaaatcctcatggtccttt 1161
Db 1140 ggagactccggtgacaaacccgggaaggtggggaacgacgtcaagtcatcatcgtgccctt 1199
QY 1162 atgtccagggccacacacgctgtcaaatgcccagatagagaggtgcgcaactcgcgaag 1221
Db 1200 acggcctgggtacacacgctgtctcaaatgggaggtacagagggcgaggggtgcgagggc 1259
QY 1222 ggagctaatctctaaaagtcgggtcccgatctcgattgggttgcgtcaactcgaccctatga 1281
Db 1260 caagcaaatcccttaaaactgttcgtagtcggttaggtgaggtcgaactcgactccatga 1319
QY 1282 agtcggaatcgtagtaatcgcggtacgacgtgcgcggtggaatacgttcccggaaccttg 1341
Db 1320 agtcggaatcgtagtaatcgcggtacgacgtgcgcggtggaatacgttcccggaaccttg 1379
QY 1342 tacacaccccgctcacacacccctgagtggggagacacccgaagtggtcttcttgaacctg 1401
Db 1380 tacacaccccgctcacacacccctgagtggggagacacccgaagtggtcttcttgaacctg 1438
QY 1402 aagagacagactactaagtgaaactcgttaaagggttggaagtcgtaacaaggtacc 1459
Db 1439 cgggagagcgttccaccaggtgtgtgttcatgactgggtggaagtcgtaacaaggtacc 1496

```

## RESULT 9

X83566  
ID X83566 standard; DNA; 1529 BP.

XX AC X83566;

XX DT 21-DEC-1999 (first entry)

XX DE 16S rDNA gene fragment from marine bacterium, isolate Wf-2.

KW Monitoring; oil; contamination; sea water; detection; flagellum;

KW Gram-negative bacterium; proteobacteria; glucose; carbon source;

KW alkane; 16S rDNA gene; ds.

XX

## Proteobacteria.

XX JP11243967-A.  
XX 14-SEP-1999.  
XX 04-MAR-1998; 98JP-0069399.  
XX 04-MAR-1998; 98JP-0069399.  
XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.  
XX WPI; 1999-564435/48.

Monitoring of oil contamination of sea water - where oil contamination is evaluated by detection of a microbe having properties from e.g. having no flagellum, being a Gram-negative bacterium, belonging to Proteobacteria, gamma subdivision, etc.z

Claim 3; Page 6; 15pp; Japanese.

The invention relates to a method for monitoring oil contamination of sea water by detecting, in the sea water, a microbe having the following properties: (1) it has no flagellum; (2) it is a Gram-negative bacterium; (3) it belongs to Proteobacteria, gamma subdivision; (4) it cannot assimilate glucose as a single carbon source; and (5) it efficiently assimilates at least one of 10-30C n-alkanes. This sequence represents a fragment of the 16S rDNA gene from the microbe of the invention, isolate Wf-2.

Sequence 1529 BP; 376 A; 352 C; 489 G; 312 T; 0 other;

Query Match 54.7%; Score 808.4; DB 20; Length 1529;  
Best Local Similarity 76.1%; Pred. No. 1.3e-241;  
Matches 1140; Conservative 0; Mismatches 316; Indels 42; Gaps 10;

```

QY 1 gatcattggtcagaactaaactgctggcgccgtcttaaacatgcaagtcgagcgggtag 60
Db 2 gatcctggtcagattgaaactgctggcgccgagcgttaaacatgcaagtcgagcgggtag 61
QY 61 caatac-----ctagcggcggaacgggtgagtaaacagtcggttaactctcc 104
Db 62 atcctagctgtctaggggcgctgcgagcggcgagtgagtaaacagtcggttaactctcc 121
QY 105 tcgagctctgggataaactttccgaaaggaaagctaaatcccgagtagctcctgttgatcac 164
Db 122 cattagtgggggataaacttgggaaacccaggctaaataaccgcataatccctacggggaa 181
QY 165 aagatttgataggtaaaagattattgcttggaagatgagcccgccgagtagctagttg 224
Db 182 agca--ggggaccttcggccttgctgtagtgatgagctcgctcggttagcttcttgg 239
QY 225 gtgaggttaattggtccaccaaggcgacgacgtgtagccgctcgagaggggtgcgcccac 284
Db 240 gtaggttaattggtccaccaaggcgacgacgtgtagctcggttagctggttagaggaatg 299
QY 285 aatggaactgagacacggtccatctactctacgggagcgagcaggttaagaatcttgcctaa 344
Db 300 accgggactgagacacggtccctactctacgggagcgagcagtgagggaatcttgacaa 359
QY 345 tgggggaaacctgaagcagcgcgcgctgtaacgaagaaggtcttcggtattgtaag- 403
Db 360 tgggggaaacctgataccagccatgcgcgctgtgtgaagaaggccttcggttgaagc 419
QY 404 ---ttcattaggcaggaaaaataagcagcaatgtg-----atgattggtaccctgcct 451
Db 420 actttcagtaggaggaaggcttatcttaatacgcgagtagtagtgcgttaccctcacag 479
QY 452 a--aacacacggtaactaactgctccagcagcgcggttaatacgtatggtgcaagcgttgt 509
Db 480 aagaagcaccggttaattctcgtgccagcagccggttaatacgaagaagggtgcgagcgttaa 539

```

```
QY 510 tcggaatcattggcgtaaaaggtgcgttaggcgggatttgtaagtccaggtgtgaaactgc 569
DB 540 tcggaatcattggcgtaaaaggtgcgttaggcgggatttgtaagtccaggtgtgaaactgc 599
QY 570 gggctcaacccggtgcctgcacttgaactcaaaactcagctcggagtttggagagcgcaagtgg 629
DB 600 gggctcaacccggtgcctgcacttgaactcagctcggagtttggagagcgcaagtgg 659
QY 630 aattccaggttagcgggtgaaatcgctagatatctggaggaaacacacagtgggcgaaggcga 689
DB 660 aatttcgggttagcgggtgaaatcgctagatatctggaggaaacacacagtgggcgaaggcgg 719
QY 690 ctgtgctgactcaaaactgacgtgagcgacgaagacgctgggttagtaaaacgggattagata 749
DB 720 ctctctgactgacattgacgtgagcggtgagaaagcggtggggagcaaacagagattagata 779
QY 750 ccccggttaactcacccctcaacgtgtctaccagttgttgggggtttta--acccctcag 807
DB 780 cctagtgtaccacggtcaacggtgtctactactgtcttgggaattcttagtattcttgg 839
QY 808 taacgaactcaacgattgaatagtagaccgctgggactatgctcgaagagtgaactca 867
DB 840 tgacgaagttaacgcgataagtagaccgctgggagtagacggcgcaaggttaaaactca 899
QY 868 aagggaattgacgggggtgcgcacacggcggtgagcatgtgttatttcgatgatacccc 927
DB 900 aatgaattgacgggggtgcgcacacggcggtgagcatgtgttatttcgatgatacccc 959
QY 928 aaaaacctcacctggcctgacatggatctga-atcaltgagagatatagacattcgg 986
DB 960 aagaacctcacggcctgacatggatctga-atcaltgagagatatagacattcgg 1019
QY 987 --gcagattcacaggtgctgaggtgtgctcagctcgtgctgagagtttgggttaa 1044
DB 1020 gagccaagtgcagaggtgctgaggtgtgctcagctcgtgctgagagtttgggttaa 1079
QY 1045 gtccgcgaacgagcgaacccctatcgatgttgc---taccttaagttgggcaactgta 1101
DB 1080 gtccgcgaacgagcgaacccctatcgatgttgc---taccttaagttgggcaactgta 1139
QY 1102 cgaactgcgggtgacaaacggaggaagcggtgagtagcgtcaaatctctatgaccttt 1161
DB 1140 gagactgcgggtgacaaacggaggaagcggtgagtagcgtcaaatctctatgaccttt 1199
QY 1162 atgtccagggccacacagtgctacaatggccgatacagaggggtgcgaactcgcgaagag 1221
DB 1200 acggcctgggtcacacagtgctacaatggccgatacagaggggtgcgaactcgcgaagag 1259
QY 1222 gagctaatctctaaagtgcgtccaggtccaggttcgagattgggtgcgaactcgcgaagag 1281
DB 1260 caagcaaatcccttaaaactgtctgtagtccggattggagttcgaactcgcgaactcga 1319
QY 1282 agtcggaatcgtagtaatcgcggatcagcatgcccgggtgagtagcgttcccggaccttg 1341
DB 1320 atcgggaatcgtagtaatcgcggatcagcatgcccgggtgagtagcgttcccggaccttg 1379
QY 1342 tacacacgcccgttcacacccactgagtggggagcaccgcaagtggttttgaacctg 1401
DB 1380 tacacacgcccgttcacacccactgagtggggagcaccgcaagtggttttgaacctg 1438
QY 1402 aaggagacagactaactaaggtgaactcgttaaaaggggtgaaagtgtaacaaagggtacc 1459
DB 1439 cgggaggacgttcacacgggttggttcaatgactgggttgaagtcgttaacaaagggtacc 1496
```

RESULT 10

X81999

ID X81999 standard; DNA; 1513 BP.

XX

AC

XX

10-SEP-1999 (first entry)

XX

```
DE B. cereus BCM 4 partial 16S rRNA gene sequence.
XX Taxane; pacilitaxel; plant; bacteria; mutation; cancer; 16S rRNA; ss.
XX Bacillus cereus.
XX W09932651-A1.
XX 01-JUL-1999.
XX 18-DEC-1998; 98WO-CA01150.
XX 22-DEC-1997; 97US-0995960.
XX (BCMB-) BCM DEV INC.
XX Bolssinot M, Gagne M, Harvey M, Helie M, Landry N;
XX Page M;
XX WPI; 1999-418940/35.
XX Production of taxane and pacilitaxel compounds
XX Disclosure; Fig 8E; 93pp; English.
XX The invention relates to a new method for the production of taxane and
CC pacilitaxel compounds that comprises culturing bacteria isolated from a
CC plant species of Taxus or bacteria produced by mutating the isolated
CC bacteria. The methods can be used for the production of taxanes such as
CC pacilitaxel, 10-deacetylcephalomannine, 7-epitaxol, 10-deacetyl-7-epi-
CC taxol, 7-epicephalomannine, baccatin III, 10-deacetyl-baccatin III,
CC cephalomannine, 7-epibaccatin III, 7-xylosyltaxol, 7-xylosyl-
CC cephalomannine, taxagifine, delta-benzoyloxy taxagifine, 9-acetyloxy
CC taxusin, 9-hydroxy taxusin, taxane Ia, taxane Ib, taxane Ic, or taxane
CC Id (claimed). The taxane and pacilitaxel compounds can be used for the
CC treatment of cancers. The methods can provide for the production of
CC taxanes and pacilitaxel at a concentration of 1-25 mu g/l.
XX
XX Sequence 1513 BP; 387 A; 341 C; 467 G; 316 T; 2 other;
```

```
Query Match 54.7%; Score 807.8; DB 20; Length 1513;
Best Local Similarity 75.1%; Pred. No. 2.1e-241.
Matches 1130; Conservative 1; Mismatches 328; Indels 46; Gaps 8;
QY 1 gatcatggctcagaactaacgctgcggcgctcttaaacatgcaagtcgagcg----- 54
DB 8 gatcctggctcagatgaacgctgcggcgctgcctaatatcatgcaagtcgagcgaaatgga 67
QY 55 -----gggtgagcaatacctagcggcgcaacgggtgagtaaacacgt-ggtaactctc 103
DB 68 ttgagagctgtctctcaagaagttagcggcgaggggtgagtaaacacgtgggtaacctgc 127
QY 104 ctccagctcggataaacttccgaagaagaaagcaagtaataaccgata-----gtcctgt 156
DB 128 ccataagcctggataaactcgcggcggttaataaccggttaataccggttaatactttggaactgc 187
QY 157 tggatcacagaatttgataggttaaaagtatttctgttgagatgagccgcggcgccattta 216
DB 188 atggttcgaattgaaaggcggtctcgtgctgctcacttatggatggaccgcgctcgattta 247
QY 217 gctagtgtgtgaggttaattgctcaccacaaaggacgacgtcgtagccgctgagaggggtgt 276
DB 248 gctagtgtgtgaggttaacggtcaccacaaaggacgacgtcgtagccgctgagaggggtga 307
QY 277 ccggccacaatggaactgagacacggtccatactctacgggagggcagcaggttaagaatc 336
DB 308 tcggccacactggactgagacacggtccacagactcctacgggagggcagcagtagggaatc 367
QY 337 ttgctcaatgggggaaacccctgaagcagcgcgcgtgaaacgaagaaggtcttcggat 396
DB 368 ttccgcaatggacgaagtcgtacgagcaacgcgcgtgagtgatgaaagggtcttcgggt 427
```

Qy	397	tgtaaagttcattgacgagaaaaataagcagcaatglgatgatggttaoctgcct-----	451
Db	428	cgtaaaactctgttattgagggaagaacaagtgtctagttgtaataagctggcacottgacgg	487
Qy	452	-----aaqcacccgctaaactacgtgccagcagccgcggtlaatactatgctgc	500
Db	488	taactaaccaagaagccacgcctaaactacgtgccagcagccgcggtlaatactatgctgc	547
Qy	501	aagcgtgttcggaatcattggcggttaaaaggtcgttagcggaatttgaagtcaggtgt	560
Db	548	aagcgttatccggaatatatttggcggttaaaagccgcgcaggtggtttcttaggtctgatgt	607
Qy	561	gaaaaactgcgggctcaaccctgacctgcacttgaataactacaagcttgaaatttggaga	620
Db	608	gaaagccacggtctaaacgtggagggctattgaaactggagacttgagtcagagaaga	667
Qy	621	ggcaagtggaaattccaggtgttagcgtgaaatcgctagatatctggagggaacaccagtgg	680
Db	668	gaaagtggaaattccatgtgtagcgtgaaatcgctagagatattggagggaacaccagtgg	727
Qy	681	cgaagcgacttgcctgcctcaaaactgcagctgtaggcgaacgaaagcgttggttagtaaacgg	740
Db	728	cgaagcgacttcttgctgttaactgcactgagggcgaaaagctggggagcaacaag	787
Qy	741	gattagatacccggttaattccacgcctaaacggttcttaacagtttggggggtttt--	798
Db	788	gattagataacctgtgtagtcacgcgcgttaaacgatgagtgctaagtttagaagggttcc	847
Qy	799	aaoccttaagtaacgaacctaacggattaaagtagacgcctcgtgggactatgctcgcagaag	858
Db	848	gccctttagtgctgaaagttaacgcattaaagcactcgcctcgtgggagtcacggccgaagc	907
Qy	859	tgaacctcaagaatttgacggggtccgcacaaagcgttgagcatgtggtttaaattcga	918
Db	908	tgaacctcaagaatttgacggggtccgcacaaagcgttgagcatgtggtttaaattcga	967
Qy	919	tgataccaccaaaacctcacctgggctgaaat--ggatctgaaatcatgtagagatatgt	977
Db	968	agcaacgcgaagaaccttaccaggtcttgacatcctctgaaaaaccttagagataggcct	1027
Qy	978	agccttcgg--gcagattcacaggtgctcatggttgcgtcagctcgtgtcgtgagatg	1035
Db	1028	ctccttcggagcagagtgacaggttgacaggttggtcatgtgttgcgtcagctcgtgtcgtgagatg	1087
Qy	1036	tgggttaagttcccgcaacgcagcgaacccct--atcgtatgttgcctaccttaagttgggc	1094
Db	1088	tgggttaagttcccgcaacgcagcgaaccccttgatcttagttgccatcattaaattgggc	1147
Qy	1095	actggtagcaaaactgcggttgacaaacccggaagaagcgggagatgaactcaaatccat	1154
Db	1148	acttaagtgactgcgggtgaacaaaccggaggaaggtggggagagcgtcaaaatcaat	1207
Qy	1155	ggccttatgttccaggggccacacagctgtacaaatggccgaacagaggggtcgccaaact	1214
Db	1208	gycccttatgaccttgggttacacacgtgtacaaatggacggtacaaaagagctgcagaacc	1267
Qy	1215	gcaagggagcgttaactctataaagtcggtcccagttcggatctgggggtctcgaactcgac	1274
Db	1268	gcgaggttggaagctaaactctataaaacccgttctcagttcgtgatttgaggtcgaactcgcc	1327
Qy	1275	ccccgaagtcgaatcgtcaglaattcgcggaatcagcatgcgcggtgtaatacgtlcccg	1334
Db	1328	tacatgaagctggaaatcgtcaglaattcgcggaatcagcatgcgcggtgtaatacgtlcccg	1387
Qy	1335	gaccttgtacacaccgcccgtcacaccaactgagttggggagcaccgcgaagtggcttttgt	1394
Db	1388	ggccttgtacacaccgcccgtcacaccaactgagttggggagcaccgcgaagtggcttttgt	1447
Qy	1395	taacggttaagagacagactactaaagttgaaactcgttaagggggtgagtcgtataacag	1454
Db	1448	aaccttttggagccagccctlaaggtgggacagatgatgtgggttgagtcgtataacag	1507
Qy	1455	gtacc	1459

```
Db 323 cactgctcagactcctacgagcgagcagtagggaataatcccgcaatggcggaagcct 382
Qy 358 gaagcagcagccgcgtgaaacgaagaaggtcttcgagttgtaaaagttcaataggcagga 417
Db 383 gacggtgcaacccgcgtgaaacgatgaaaggttttcgagtcgtaaaagtctgttagagg 442
Qy 418 aaaaataagcagcaatgatgatgttaactgctc-----aaagcacccg 461
Db 443 aagaacaagtcgcttgaataaggttcgacacctgacggtacctcaacagaaagcccccgg 502
Qy 462 ctcaactacgtgccagcagccgcgtgaaatacgtatggtcgaagcttcttcggaatacattg 521
Db 503 ctcaactacgtgccagcagccgcgtgaaatacgtatggtcgaagcttcttcggaataattg 562
Qy 522 ggcgtaaaagggctgtaggcggaattgttaagtcaggtgtgaaatacgtgggtcacaaccg 581
Db 563 ggcgtaaaagggctgtaggcggaattgttaagtcaggtgtgaaatacgtgggtcacaaccg 622
Qy 582 tggcctgcaactgaaactacaagctgaggttggagagcgaagtgaaattccaggtgt 641
Db 623 ggaaggtcattggaacacgggggaactgagtgtaggaggaagtgaaattccacggtg 682
Qy 642 agcgtgtaaatcgttagatatctgaggaacaccagtggtggaagcgacttctggtctca 701
Db 683 agcgtgtaaatcgttagatatctgaggaacaccagtggtggaagcgacttctggtctca 742
Qy 702 aaactgacgtgagcagcagaaacgctgggttagtaaacgggtagatacccccggttaacc 761
Db 743 caactgacgtgagcagcagaaacgctgggttagtaaacgggtagatacccccggttagtcc 802
Qy 762 acgcctaaacgttcttaccagttgttgggggttt--taaccctcagtaacgaacacaa 819
Db 803 acgcgttaaacgatgagtgctagggttgggttgggttttgcatacccttagtgcggaagttaa 862
Qy 820 cggattagtagacgcctcctgggaactatgctgcgaagtagtgaactcgaaggaattgacg 879
Db 863 cacattagcactcgcctcctgggaactatgctgcgaagtagtgaactcgaaggaattgacg 922
Qy 880 ggggtccgacaaagcgtggaacatggtgttaattcagatgatacccccaaaacacacc 939
Db 923 ggggtccgacaaagcgtggaacatggtgttaattcagatgatacccccaaaacacacc 982
Qy 940 tgggtctgaca-----tggatctgaatcatgtagagatatagccttcgggcagattc 994
Db 983 aggtttgacatcctctgacacctctggagacagagcgttcccttcgggggacagagtg 1042
Qy 995 acagatgtctgactgtctcagctcgtgctgagatggttggttaagttcccgcaac 1054
Db 1043 acaggtggtgcatggtgtcgtcagctcgtgctgagatggttggttaagttcccgcaac 1102
Qy 1055 gaggcacaacccct-atcgatgttctacacttaagttgggcaactggtgcgaactgcgcg 1113
Db 1103 gaggcacaacccctgacttagttgcagcattcagttgggcaactggtgcgaactgcgcg 1162
Qy 1114 tgacaacccgaggaagcgggagtagcgtcaaatccatgacctttagtccagggcc 1173
Db 1163 tgataaacccggaaggtggggaagcgtcaaatccatgacctttagtccagggcc 1222
Qy 1174 acacacgtgctacaaatggcgcatacagaggggtcgcccaactcgcaagagggagactatctc 1233
Db 1223 acacacgtgctacaaatgggtatcaaaagggcagcgagacccgaggttaagcgaatccc 1282
Qy 1234 taaaagtcggtcccaagttcggattgggtctcgaactcgaccccatgaagtcgggaatcgc 1293
Db 1283 ataaagccattctcagttcggattcgggtcgaactcgctgcaagcgggaattgc 1342
Qy 1294 tagtaatcgcgatcagatgcgcggtggaatacgttcccgacctgttacacacgcgcc 1353
Db 1343 tagtaatcgcgatcagatgcgcggtggaatacgttcccgacctgtgttacacacgcgcc 1402
Qy 1354 gtcacacacacgtggtgggagcaccacgaagtgggtcttctgttaacgtgaagcagacac 1413
Db 1403 gtcacacacacagagttgttaacacccgaagtcg-gtgcggtaaccttcttgagcagacc 1461
```

```
Qy 1414 tactaagtgaaactcgttaaaaggggtgaagtcgtaacaaggtacc 1459
Db 1462 gncgaaggtggacacagatgttggggtgaagtcgtaacaaggtacc 1507
```

## RESULT 12

T01866

ID T01866 standard; DNA; 1526 BP.

XX

AC T01866;

XX

DT 03-AUG-1999 (first entry)

XX

DE P.cepacia 16S rRNA gene sequence.

XX

KW 16S rRNA; KK01; primer; PCR; amplification; probe; hybridisation; detection; diagnosis; ds.

XX

OS Pseudomonas cepacia.

XX

PN JP07255486-A.

XX

PD 09-OCT-1995.

XX

PF 23-MAR-1994; 94JP-0051739.

XX

PR 23-MAR-1994; 94JP-0051739.

XX

PA (CANO ) CANON KK.

XX

DR WPI; 1995-378541/49.

XX

PT Pseudomonas cepacia KK01 strain 16S rRNA gene - also related probes and primers, useful for specific detection of P.cepacia strain KK01

XX

PS Claim 1; Page 21; 21pp; Japanese.

XX

CC This sequence represents the 16S rRNA gene of Pseudomonas cepacia strain KK01. Fragments of the nucleic acid sequence (see T01872-T02316) are useful as primers and probes for the specific detection of P.cepacia strain KK01.

XX

SQ Sequence 1526 BP; 382 A; 352 C; 486 G; 306 T; 0 other;

Query Match 54.5%; Score 805.4; DB 16; Length 1526;

Best Local Similarity. 75.3%; Pred. No. 1.2e-240;

Matches 1124; Conservative 0; Mismatches 326; Indels 43; Gaps 8;

Qy 2 atcatggtctagaactaaacgtgcgcgcgtctttaaaccatgcaagtcgagcgggtagc 61

Db 9 atcctggtcagattgaacgctggcgcatgcttacacatgcaagtcgagcggcagcac 68

Qy 62 aataactagc-----ggcgaacgggtgagtaaacacgtggtaattctctcc 107

Db 69 ggggtctgcaacctgggtggcgagtggtggaacgggtgagtaaacacgtggtaattctctcc 128

Qy 108 ggtctgggataaacttccgaaagaaagctaataccggatagctctgttgatcacaaag 167

Db 129 tagtggggatagccgcggaagcggatttaaccgcatacagatcttcagat--gaaa 186

Qy 168 atttgataggtaaaagtattctgttgagatgagccccgcggccgattagttggtg 227

Db 187 gcgggggaaccttcggcctcgctatagggttggcggatggctgattagttggtg 246

Qy 228 aggtaatggtcaccaagcgacagctggtacccgctgagaggtgtcccgccacaat 287

Db 247 ggggtaaaagccaccacaagggcgacatgactggtgtgagagcagaccagccacct 306

Qy 288 ggaactgagacacggtcccatctctacgggagggcagcagttlaagaattctgtcaatgg 347

Db 307 gggactgagacacgcccacagactcctacgggagggcagcagttggggaattttggacaatgg 366







```
Db 240 gggggtaagggccacaaagcgacgacccgtaactggctgagagatggccaggtcac 299
Qy 285 aatgaactgagacacggtccactactctacgggagcgagcaggttaagaacttgcctcaa 344
Db 300 accgggactgagacgacggccggaactctacgggagcgagcagtggaactcttgacaa 359
Qy 345 tggggaaacccctgaagcagcagcccgctgtaaacgagaaggtctcggaatgttaagt 404
Db 360 tgggcgaagcctgacacgacccatccgctgtgtgaaagaaggcctcgggttgtaaaagc 419
Qy 405 tcattaggcaggaataa-----agcagcaatgtgatgtgacctgctc 451
Db 420 eetttcagtaggagaaggccttgggttaataccctcgagtaacttgacgttacctacag 479
Qy 452 a--aagccggctgaactactgcccagcagccgctgtaatacgtatggtgcaagcgtgt 509
Db 480 aagaagcagcggttaattctgtgccagcagccgctgtaatacgaaggtgcgagcgttaa 539
Qy 510 tcggaatcattggcgctgaagggtcgtaggcggatttgaagtcaggtgtgaaactgc 569
Db 540 tcggaatcattggcgctgaaggcgcgtagcggtgtgttaagtcgagtgtaaaagccca 599
Qy 570 gggctcaaccgctggcctgcactgaaactacaagctcctggagtttggagagcgaagtgg 629
Db 600 gggctcaaccctgggaattgcatccgatactggcagcgttagagcagtagaggggctgg 659
Qy 630 aatccaggttagcgggtgaatgcgtgatatctggagggaacaccagtggcgaaggcga 689
Db 660 aattccgggttagcgggtgaatgcgtgagatcgtagatcggaagaaacaccagtggcgaaggcgg 719
Qy 690 ctgtgtgctcaaaactgaagcgtgaggcaacgaagcgtgggttagtaaaacgggattagata 749
Db 720 cctcctgactgacactgacgtcgaggtgagcaagcgtggggagcaaacagagattagata 779
Qy 750 ccccggttaatcacccctcaaacgtgtgtaccaggtgttgggggttta--accctcag 807
Db 780 cctggttagtcacccgttaaacgattgtctactagccttgggggtccttagtgactttgg 839
Qy 808 taacgaactaacgattgaatagtagccgctgggactatgctcgaagagtgaaactca 867
Db 840 tggcgagctaacgcgataagtagaccgctgggagtagcggcgcaaggtttaaactca 899
Qy 868 aaggaaatgacgggtgcgcacacgctgagagcagtggttgaatttcgatgatacccc 927
Db 900 aatgaattgacggggggccgacagcgtggagcatgtggtttaaattcgatgcacgcg 959
Qy 928 aaaaacctcacctggccttgacatggaattga-atcattagagatatatgagccttcgg 986
Db 960 aagaaccttacacggccttgacatcctcgaaacttctagagatagattggcgttcgg 1019
Qy 987 --gcagattcacagtgctgcatgattgtgctcagctcgtgctgagatgttggtgttaa 1044
Db 1020 gagcgagtgacaggtgctgcagtggctgctcagctcgtgctgagatgttggtgttaa 1079
Qy 1045 gtccgcgaacgagcgaacccctatcgtatgttgta--ccttaagttgggcactgggtac 1102
Db 1080 gtccgcgaacgagcgaacccctgtccttagttgcagacacttcgggtgggaactctagg 1139
Qy 1103 gaaactgcgggtgacaaacggaggaagcggtgagtcgaactcaatctcatggccttta 1162
Db 1140 gggactgcgggtgacaaacggaggaagcggtggagcagcgtcaagtcatacgtggcctta 1199
Qy 1163 tgtccaggggccacacagtgctacaaatggccga tacagagggtcgccaaactcgcaagag 1222
Db 1200 cggcctgggtacacagtgctacaaatgggttggtacagaggttgcgaagtcgcgagcg 1259
Qy 1223 gagcctaactctaaagtcggtcccaagttcggatgggggtcgtgcgaactcgaccocatgaa 1282
Db 1260 gagcctaactctcaaaagcgaatcgtagtcggatgggttcgtgcgaactcgactccatgaa 1319
Qy 1283 gtcggaatcgttagtaatgcggatcagcatcccggtggaatagcttccccgacctgt 1342
Db 1342 gtcggaatcgttagtaatgcggatcagcatcccggtggaatagcttccccgacctgt 1342
```

```
Db 1320 gtcggaatcgttagtaatgcggatcagaatccgcggtggaatcgttccccgggacctgt 1379
Qy 1343 acacaccgcccgtcacaccacctgagtggtgggagcaccgaaagtgtcttgttaaccgta 1402
Db 1380 acacaccgcccgtcacaccacctgagtggtgggagcaccgaaagt-tagtctaaccttc 1438
Qy 1403 aggaacagcagactactaaagtgaaactcgttaaagggtgaaagtcgttaacaaaggtacc 1459
Db 1439 gggagagcagattaccacggtgtggttcactgactggtggaagtcgtaacaaaggtagc 1495

RESULT 14
X83568
ID X83568 standard; DNA; 1535 BP.
XX
AC X83568;
XX
DT 21-DEC-1999 (first entry)
XX
DE 16S rDNA gene fragment from marine bacterium, isolate K2-1.
XX
KW Monitoring; oil; contamination; sea water; detection; flagellum;
KW Gram-negative bacterium; Proteobacteria; glucose; carbon source;
KW alkane; 16S rDNA gene; ds.
XX
OS Proteobacteria.
XX
PN JP11243967-A.
XX
PD 14-SEP-1999.
XX
PF 04-MAR-1998; 98JP-0069399.
XX
PR 04-MAR-1998; 98JP-0069399.
XX
PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX
DR WPI; 1999-564435/48.
XX
PT Monitoring of oil contamination of sea water - where oil contamination
PT is evaluated by detection of a microbe having properties from e.g.
PT having no flagellum, being a Gram-negative bacterium, belonging to
PT Proteobacteria, gamma subdivision, etc.z
XX
PS Claim 3; Page 7; 15pp; Japanese.
XX
CC The invention relates to a method for monitoring oil contamination of
CC sea water by detecting, in the sea water, a microbe having the following
CC properties: (1) it has no flagellum; (2) it is a Gram-negative bacterium;
CC (3) it belongs to Proteobacteria, gamma subdivision; (4) it cannot
CC assimilate glucose as a single carbon source; and (5) it efficiently
CC assimilates at least one of 10-30C n-alkanes. This sequence represents
CC a fragment of the 16S rDNA gene from the microbe of the invention,
CC isolate K2-1.
XX
SQ Sequence 1535 BP; 376 A; 349 C; 494 G; 316 T; 0 other;
```

Query Match 54.5%; Score 805; DB 20; Length 1535;  
Best Local Similarity 75.6%; Pred. No. 1.5e-240;  
Matches 1131; Conservative 0; Mismatches 325; Indels 41; Gaps 9;

```
Qy 1 gatcgtgctcagaactaacgctggcgcgctctttaaactgaagtcgagcggg--- 57
Db 9 gatcctggtcagattgaacgctggcgcgagcgccctaacaacatgcaagtcgagcggtaa 68
Qy 58 -----tagcaataactagcgcgagcggtgagtaaacacgctggttaattccc 104
Db 69 ggagaagctgtctctcgtcagcgcgagcggtgagtaaacgctgagaatctgcc 128
Qy 105 tccgagctcgggataacttccgaaaggaaagcgaataacccgagatcctgttgatcac 164
Db 129 cattgtg9gggataactcgggaaactcaagcgaataacccgagcgggggaa 188
```



```
OY 1 gatacgtgctcagaactaacgctgcggc-gcgttattaaacatgcaagtcgagcgg---- 55
Db 15 gaucucgucagauugacgucgucgcaugcuuuaacacagcaagcaacgguaac 74
OY 56 -----ggtagcaatcacctagcggcgacggtgagtaacacgtggttaattctctccga 109
Db 75 agguucucggaugcugacgaguggcgaacgcgguaguaaacaucggaacgucgucua 134
OY 110 gtcctgggataacttcccgaaggaaagtaataaccggatagtcctgttggatcacaaagt 169
Db 135 gugggggaaauaacuacgaaagagucuaauacgcgaugagauacucggaugaagca- 193
OY 170 ttgataagtaaaagtatttctgttgagatgagcccgcgccgattagctagttggtgag 229
Db 194 -gggaccuucggccuugucuaagagcgguagugcgagauuagguuuguggg 252
OY 230 gtaattggtcaccaaggcgacgctggtagccgctgagaggtgtccggccacaaagg 289
Db 253 guaaaggcuuaccagcugcgauucugucugugagagagacgaccagccacacug 312
OY 290 aactgagacacggtccatctactcctcagggagcgacagttaaagaatcttgcctcaatgggg 349
Db 313 gacugagacacggcccagacucucagcgaggcgagcaguggggaauuuggacaaugggc 372
OY 350 gaaaccctgaagcagcgcgcgtgaacaaagaagcttctcgatgttaagtctatt 409
Db 373 gaaagcugacagcgaugcccgugcaggaaggcgccuucggguuuaaacucugcuu 432
OY 410 aggcaggaaaaataagcag-----caatgtgatgtggtacctgccta--aa 454
Db 433 uguacggaacgaaagccugggcuauauccccgggucaugacgguacgguagaa 492
OY 455 gcaacggctaaactacgtgccagcagccggttaatacgtatggtcgaagcgtgttcgga 514
Db 493 gaccggcuuacucugccagcagccgguuauacugagggugcaagcguaauaucgga 552
OY 515 atcattgggctaaagggtgcgtagcggaatttgaagtcaggtgtgaaacctgcggct 574
Db 553 auuacgggcuuagagcgucgagcgggguuuguaacaguguguaaucccccgggcu 612
OY 575 caaccctgtgctgacttgaactacaaagtcgtgaggttgggagaggcaagtgaattc 634
Db 613 caaccugggaaacugccaauugacugcgaaggcuagugcgcgagggggaugaauc 672
OY 635 cagggtgtagcgggtgaatgcgtagatatactggagaaacacacagtcggcgaaggcactgc 694
Db 673 cgcugugagcagugaaugcguuauugcggaugcggaacacccgaugcgaaagcauuccc 732
OY 695 tggctcaaaactgacgtgaggcagcgaagcgtggttagtaaaacgggattagataccccg 754
Db 733 ugggccugcacugacgcucaugcagaaagcugugggagcaaacaggauuagauaccucg 792
OY 755 gtaatccacgcctaaacgttgtaccagttgttgggggttttaacctcagtaacgaa 814
Db 793 guaguccacgcgccuaaagcuaugcuacugguuugguuuaacugacucaagaaacgaa 852
OY 815 cctaacgcattaagtagaccgcctgggactatgctcgaagagtcgaactcaaaaggaat 874
Db 853 gcuaaacgcgugauguaugaccgcugggagucggccgcaagguuagaaacuaaaggaa 912
OY 875 tgacgggtccgcacaagcgtgagcagctgtgttattcgtatgatacccccacaaacc 934
Db 913 ugacggggaccgcacagcggugaugaugugguuuaauucugcaacgcgaaaaacc 972
OY 935 tcacctgggctgacatggatcatcatgtagagatatagaccttcgggcaga--- 991
Db 973 uuaccacuuuagacauggcaggaacuuaccagagaugguuugugcugaaagagaacc 1032
OY 992 ---ttcacagggtgcataggttgtcgtacgctcgtcgtagatggtgggttaagtccc 1049
Db 1033 ugacacagugcugcaguggcugucucagcagcugcugugagugaugugguuuaagucc 1092
OY 1050 gcaacgagcgcaaacccctatcgtatgttgcctacctaagtgtgggcactggtacgaaactg 1109
```

```
Db 1093 gcaacgagcgcgaacccuugccaauugugcuacauucaguuagcacucuaaugggacug 1152
OY 1110 ccggtgacaaaaccggagggaagcggggtagcgtcaaatcctctatggcctttatgtccag 1169
Db 1153 ccggugacaaaaccggagggaagcgggggaugacugcaaguccuauuagguug 1212
OY 1170 gcccacacgctgctacataatggccgatacagagggtcgccaactcgcaagggagctaa 1229
Db 1213 ggcuacacacgcuacuaaauugcuguaacaaaggguugccaaaccgcgaggggagcuua 1272
OY 1230 tctctaaagtcgggtcccagttcggattcgggtcgtcgaactcgaccccatgaagtcgaa 1289
Db 1273 ucccauaagccagucugucggaugccgagucgagucgacucgacugcgugagucgaa 1332
OY 1290 tgcgtagtaatacgcggatcagcatcgcggtgaatacgttccccgaccttgcacaccc 1349
Db 1333 ucgcuaaauucguggaucagaaucaucgguagaauucucccgguucuguaacaccc 1392
OY 1350 gccctcacacacacccctgagtggggagcaccggcgggtcttcttgaacgcgtaagagac 1409
Db 1393 gcccgucacacacaugggagcgguucucgcagaaugag-guagccuaaccguuaggaggg 1451
OY 1410 agactactaaagtgaaactcgtaaaagggtgaagtcgttaacaaggtacc 1459
Db 1452 cgcuuaccacgcggggguucugagcggggugagucgguuacaaagagc 1501
```

Search completed: March 27, 2001, 08:22:56

Job time: 5298 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March. 27, 2001, 06:53:33 ; Search time 132.3 Seconds  
(without alignments)  
1799.197 Million cell updates/sec

Title: US-09-380-826A-1

Perfect score: 1477

Sequence: 1 gatcatggctcagaactaac.....cagtaaatgattcctgcag 1477

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A-COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B-COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6-COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/PCTUS-COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	814.2	55.1	1536	2 US-08-642-229A-1	Sequence 1, Appli
2	805.4	54.5	1521	3 US-08-501-126-20	Sequence 20, Appl
3	803.8	54.4	1536	1 US-08-114-695A-7	Sequence 7, Appli
4	801.8	54.3	1512	3 US-08-995-960-2	Sequence 2, Appli
5	801.6	54.3	1555	2 US-08-757-653-160	Sequence 160, App
6	792.4	53.6	1517	3 US-09-248-528-2	Sequence 2, Appli
7	789.8	53.5	1516	3 US-09-248-528-3	Sequence 3, Appli
8	780.2	52.8	1542	1 US-08-114-695A-1	Sequence 1, Appli
9	780	52.8	1542	2 US-08-757-653-158	Sequence 158, App
10	779.8	52.8	1474	1 US-08-114-695A-8	Sequence 8, Appli
11	778.4	52.7	1556	3 US-08-995-960-1	Sequence 1, Appli
12	772.6	52.3	1484	2 US-08-632-470-53	Sequence 53, Appl
13	770.6	52.2	1518	1 US-08-114-695A-6	Sequence 6, Appli
14	768.4	52.0	1485	1 US-08-299-810A-27	Sequence 27, Appl
15	766.6	51.9	1542	2 US-08-875-445-21	Sequence 21, Appl
16	754.6	51.1	1494	2 US-08-632-470-49	Sequence 49, Appl
17	754.2	51.1	1452	2 US-08-642-229A-2	Sequence 2, Appli
18	753.6	51.0	1455	2 US-08-642-229A-3	Sequence 3, Appli
19	749.6	50.8	1464	2 US-08-938-858-1	Sequence 1, Appli
20	749	50.7	1508	2 US-08-632-470-44	Sequence 44, Appl
21	747.8	50.6	1452	1 US-08-276-943-1	Sequence 1, Appli
22	747.8	50.6	1452	2 US-08-716-841-1	Sequence 1, Appli
23	744.6	50.4	1430	2 US-08-902-518A-1	Sequence 1, Appli
24	743.8	50.4	1415	2 US-08-632-470-52	Sequence 52, Appl
25	741.8	50.2	1440	2 US-08-632-470-26	Sequence 26, Appl
26	741	50.2	1462	3 US-09-191-099-3	Sequence 3, Appli
27	740.2	50.1	1440	2 US-08-632-470-33	Sequence 33, Appl
28	740.2	50.1	1440	2 US-08-632-470-35	Sequence 35, Appl

#### ALIGNMENTS

##### RESULT 1

US-08-642-229A-1

; Sequence 1, Application US/08642229A

; Patent No. 5874291

; GENERAL INFORMATION:

; APPLICANT: Herwig, Russell P.

; APPLICANT: Bielefeldt, Angela R.

; APPLICANT: Stensel, H. David

; APPLICANT: Strand, Stuart E.

; TITLE OF INVENTION: Degradation of Environmental Toxins by a

; TITLE OF INVENTION: Filamentous Bacterium

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Christensen O'Connor Johnson and Kindness PLLC

; STREET: 1420 Fifth Avenue, Suite 2800

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: WA 98101-2347

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/642,229A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/246,865

; FILING DATE: 20-MAY-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Sheiness, Diana K.

; REGISTRATION NUMBER: 35,356

; REFERENCE/DOCKET NUMBER: UOFW19233

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 682 8100

; TELEFAX: (206) 224 0779

; TELEX: 4938023

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1536 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: nucleic acid

; DESCRIPTION: "16S ribosomal DNA"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Comamonas testosteroni ATCC No. 5874291 11996

Sequence 41, Appl  
Sequence 24, Appl  
Sequence 29, Appl  
Sequence 45, Appl  
Sequence 34, Appl  
Sequence 47, Appl  
Sequence 23, Appl  
Sequence 36, Appl  
Sequence 37, Appl  
Sequence 46, Appl  
Sequence 30, Appl  
Sequence 31, Appl  
Sequence 28, Appl  
Sequence 43, Appl  
Sequence 8, Appl  
Sequence 25, Appl  
Sequence 38, Appl

US-08-642-229A-1

Query Match 55.1%; Score 814.2; DB 2; Length 1536;  
Best Local Similarity 74.7%; Pred. No. 4.8e-284;  
Matches 1113; Conservative 0; Mismatches 343; Indels 33;

Qy	1	gataatggctcagaactaacacgctg	ggcgcgctcttaaacaatgcaagtc	cgagcggggtag	60	
Db	16	gATCTGGCTCAGATTAAGACGCT	GGGGCATGCTTTTACATGCAAGT	CGAACGGTAAGA	75	
Qy	61	caa-----tac	tagcgcgaacgggtgagtaaac	gctggtaactctctccgag	110	
Db	76	GGTCTTCGGATGCTGACGAGT	GCGCAACGGGTGAGTAATACAT	CGGACGTCGCTAGTAG	135	
Qy	111	tcctgggataaacttc	cgaaaggaaagctaataccgga	tatgctctgttgatcacaaagatt	170	
Db	136	TGGGGGATAACTACTCTGAAAGAG	TAGCTAATACCGCATGAGATCT	ACGGGATGAAGAACA--	193	
Qy	171	tga	tagataaagatttatctctggagat	gagccgcgccgattagcttagttggtgaag	230	
Db	194	GGGACCTTCGGGCTTGCTACT	AGACGGCTGATGGCAGATTAAG	TAGTTTGGTGGGG	253	
Qy	231	taa	tggtctcacaaggcgaacgac	tcggtagcggcgctgagagggtg	ctccggccacaatgga	290
Db	254	TAAAGGCTTACCAAGCCTGCGAT	CTGTAGCTGTCTGAGAGGAC	GACACGACACATGGG	313	
Qy	291	actgagacacggtctc	attctctcctcgggagcgagctta	agaatcttctcgaatggggg	350	
Db	314	ACTGAGACACGGCCAGAC	TCCCTACGGGAGGCAAGTAGTGGG	GAATTTTGACAAATGGGGG	373	
Qy	351	aaaccc	tgaaagcagcgcgcgctgaa	cgaagaaggtcttcgga	tgtgtaaagttcatla	410
Db	374	AAAGCCTGATCCAGCAAT	CGCGCTGCAGGATCAAGGCCCT	CGGGTTGTAAACTGCTTTT	433	
Qy	411	ggcaggaaaaataagca-----	-----gcaatg	tgatgacgttacctgccta--	aaag	455
Db	434	GTACGGCAACGAAGAAGC	TGTGGGGCTAATATCCCCGGGT	CTATGACGGGTACCGGTAGAAATAAG	493	
Qy	456	cacggctaaactacg	tgcacgacgcgcggtta	atagctatggtgcgaagctgtgtcgcga	515	
Db	494	CACGGGCTAACTACGT	AGCTGCCAGCAGCCGGTAA	TACGTAGGTTGCAAGCGTTAATCGGAA	553	
Qy	516	tcat	tgggcgtaaaagggtgcg	taggcggatttctaagctcaggtgtgtaaaactgcggctc	575	
Db	554	TTACTGGGCGTTAAAGCT	GTGCGCAGCGGCTTTTGTAA	GACAGTGTGCTGAAATCCCGCGGCTC	613	
Qy	576	aacc	gtggctcacacttgaaact	caaaactcgaagcttgaggttgggagagcgaagtgaattcc	635	
Db	614	AACCTGGGAAC	TGCACTTGCTGCAAGGCT	AGAGTGGCGAGAGGGGGATGGAATTC	673	
Qy	636	agg	tgtagcgggtgaaatgcg	tagatactctggaggaacacacag	tggcgaaggcgacttgt	695
Db	674	CGGTGTAGCAGTGAAT	TGCGTGTAGATATGCG	GAGGAACACCGATGCGCAGGCAATCCCT	733	
Qy	696	ggct	caaaactgcctgaggaacg	aaagcgtgggtagtaaacagggaatagatacccccgg	755	
Db	734	GGGCTGCATGACGCTCAT	GTGCAAAAGCGTGGGAGCAAA	CAGAGTTAGATACCCCTGG	793	
Qy	756	taatcc	acgcctcaaacg	tgtctaccagtgtgtgggggttttaac	ccctcgaataacgaac	815
Db	794	TAGTCCAGCCCTTAAC	GATGTCACTGTTTGGGCTT	TAACTGACTGACTGATGAACGAAG	853	
Qy	816	ctaacg	gattaa	gtagacgcctcggggactatctcgcgaagtagtaaacctcaaaagatt	875	
Db	854	CTAACGCGTGAAGTT	ACC	GGCTTGGGAGGTACCGCGCAAGGTTGAAC	ACTCAAAAGGAAT	913
Qy	876	gacg	gggtccgcacaa	cgctgagacatgatttaattcaatgataccccaacact	935	
Db	914	GACGGGACCCGACAA	AGCCGTGATGTGGTTTAA	TTCGATGCAACCGCGAAACCT	973	
Qy	936	caact	gggtctgacatggatctg	aaatcatgtagagataatatgac	gacctctgggcaga----	991

Db	863	CACATTAAAGCACTCCGCTTGGGAGATAGCGCGCAAGGCTGAAACTCAAGGAATTGACG	932
Qy	880	ggggtccgcacaagcgggtggagcatgtgttttaattcgatgataaccaccaaaacccctcaacc	939
Db	923	GGGGCCGCGACACAGCAGTGGACCATGTGTTTAAATTCGAAGCACACGACGAGACCTTACC	982
Qy	940	tgggcttgaca-----tggatcgtgaatcatgtagagatatatagagccttcgggcagattc	994
Db	983	AGGTCTTTGACATCTCTGACACCTCTGGAGACAGAGCGTTCCCTTCGGGGGACAGAGTG	1042
Qy	995	acaggctgcacatggttgcgcagactcgctcgtagagatgttggtttaagttccgcgaac	1054
Db	1043	ACAGGTGGTGCATGGTGTGTCAGCTCGTTCGCGAGATGTTGGGTAAAGTCCCGCAAC	1102
Qy	1055	gagcgcaacccct-atcgatatgtctaccctaaagttgggcactgggtacgaaactgcg	1113
Db	1103	GAGCGCAACCCCTTGATCTTAGTTGCGCAGCATTCAGTTGGGCACTCTTAAGGTGACTGCCGG	1162
Qy	1114	tgacaacccggaggagcggggatgacgtcaaatccctcatggcctttatgtccagggcc	1173
Db	1163	TGATAAACCGGAGGAAGTGGGATCAGCTCAAAATCATCATGCTCCCTTATGACCTGGGCT	1222
Qy	1174	acacagtgctacaatggccgatacagaggggtcgccaactcgcaagaggagagtaacttc	1233
Db	1223	ACACAGCTGCTACATGGATGGTGTACAAAGGCGACGACCGCAGGTTAAGCGCAATCCC	1282
Qy	1234	taaaagtcggtccaggttcggattggggtcctgcgaactcgaccatgaagtcggaatcgc	1293
Db	1283	ATAAAGCCATTCTCAGTTCGGATTTCAGGCTGCACACTGCCCTGCATGAAGCCGGAATTGC	1342
Qy	1294	tagtaatcggatcagcatgcgcggtggaatacgttccccggacctgtcacacacgccc	1353
Db	1343	TAGTAATCGGGATCAGCATGCCGGGTGAATACAGTTCCTCCGGGTCTTGTACACACCGCC	1402
Qy	1354	gtcacaccctgagtggggagacacccgaagtggtcttgttaaccgttaaggagagacag	1413
Db	1403	GTCAACACACGAGAGTTGTATACACCCGAAGTCG-GTCGGTGAACCTTTTGGAGCCAGCC	1461
Qy	1414	tactaaggtgaaactcgtaagggggtgaagtcgtaacaaaggtacc	1459
Db	1462	GNGCAGGTGGACAGATGATTGGGGTGAAGTCGTAAACAGGTATC	1507
RESULT 3			
US-08-114-695A-7			
; Sequence 7, Application US/08114695A			
; Patent No. 5508193			
; GENERAL INFORMATION:			
; APPLICANT: Mandelbaum, Raphael T.			
; APPLICANT: Wackett, Lawrence P.			
; TITLE OF INVENTION: DEGRADATION OF S-TRIAZINES IN SOIL AND			
; TITLE OF INVENTION: WATER			
; NUMBER OF SEQUENCES: 8			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: SCHWEGMAN, LUNDBERG & WOESSNER, P.A.			
; STREET: 3500 IDS CENTER			
; CITY: MINNEAPOLIS			
; STATE: MN			
; COUNTRY: USA			
; ZIP: 55402			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/114,695A			
; FILING DATE: 31-AUG-1993			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: MUEITING, ANN M.			
; REGISTRATION NUMBER: 33,977			









```

: GENERAL INFORMATION:
: APPLICANT: Oriel, Patrick J
: APPLICANT: Padmakumar, Rugmini
: APPLICANT: Kim, Sang H
: TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
: TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
: FILE REFERENCE: MSU 4.1-401
: CURRENT APPLICATION NUMBER: US/09/248,528C
: CURRENT FILING DATE: 1999-02-10
: EARLIER APPLICATION NUMBER: 60/083,485
: EARLIER FILING DATE: 1998-04-29
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 1517
: TYPE: DNA
: ORGANISM: Bacillus sp.
: FEATURE:
: NAME/KEY: rRNA
: LOCATION: (1)..(1517)
: PUBLICATION INFORMATION:
: DATABASE ACCESSION NUMBER: 226929
: DATABASE ENTRY DATE: 1998-07-02
: US-09-248-528-2

```

[illegible]

```

RESULT* 7
US-09-248-528-3
; Sequence 3, Application US/09248528C
; Patent No. 6153415
; GENERAL INFORMATION:
; APPLICANT: Oriel, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
; TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
; FILE REFERENCE: MSU 4.1-401

```

;; CURRENT APPLICATION NUMBER: US/09/248,528C  
;; CURRENT FILING DATE: 1999-02-10  
;; EARLIER APPLICATION NUMBER: 60/083,485  
;; EARLIER FILING DATE: 1998-04-29  
;; NUMBER OF SEQ ID NOS: 18  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 3  
;; LENGTH: 1516  
;; TYPE: DNA  
;; ORGANISM: Bacillus pallidus  
;; FEATURE:  
;; NAME/KEY: rRNA  
;; LOCATION: (1)..(1516)  
;; PUBLICATION INFORMATION:  
;; DATABASE ACCESSION NUMBER: 226930/GenBank  
;; DATABASE ENTRY DATE: 1997-05-14  
US-09-248-528-3

Query Match 53.5%; Score 789.8; DB 3; Length 1516;  
Best Local Similarity 75.0%; Pred. No. 3.2e-275;  
Matches 1124; Conservative 0; Mismatches 327; Indels 48; Gaps 9;  
  
QY 9 CTCAGAACTAACGCTGGCGCGCTCTTAAACATCAAGTCGAGCGGGTAGCAATACCT 68  
DB 1 CTCAGGAGCAAGCTGGCGCGCTCTTAAACATCAAGTCGAGCGGGTAGCAATACCT 60  
  
QY 69 AG-----CGGCAACGGGTGAGTAACACCTGGTAATCTTCTCCG-AGTCT 113  
DB 61 TGCTCTTTAGTTAACGCGGACGGGTGAGTAACACCTGGGCAACCTGCCCTGCAGACT 120  
  
QY 114 GGGATAAATTCGAAAGAAAGCTAAATACCGGATAGCTCTGTGGATCAACAAGATTGA 173  
DB 121 GGGATAAATTCGAAAGAAAGCTAAATACCGGATAGCTCTGTGGATCAACAAGATTGA 180  
  
QY 174 TAGGTAAG-----ATTATTGCTGGAGATGAGCGCGCGGCGGATAGTAGTGG 225  
DB 181 GTGAAAGCGGCTTTAGCTGCTCACTCAGGATGGCGCGGCGGATAGTAGTGG 240  
  
QY 226 TGAGGTAATGCTCAACAGCGACGATCGTAGCGGCGCTGAGAGGCTGTCGGGCCACA 285  
DB 241 TGAGGTAATGCTCAACAGCGACGATCGTAGCGGCGCTGAGAGGCTGTCGGGCCACA 300  
  
QY 286 ATGGAACTGAGACACGGTCCATACCTCAGGAGGAGCGAGTAAAGATCTGCTCAAT 345  
DB 301 CTGGGACTGAGACACGGTCCATACCTCAGGAGGAGCGAGTAAAGATCTGCTCAAT 360  
  
QY 346 GGGGAAACCTTGAAAGCGACGCGCTGAGCAAGAAAGTCTTGGATTGTAAGTT 405  
DB 361 GAGCAAGATCTGACGGAGCAACGCGCGTGGAGCAAGAAAGTCTTGGATTGTAAGTT 420  
  
QY 406 CATTAGCGAGGAAATAAGCAGCAATGTA-----TGATGTAACCTGCT 451  
DB 421 CTGTGTGAGGAGAAACAAGTCGCTCGAACCGGCGGTACCTTGACGTAACCTGACG 480  
  
QY 452 A--AAGCCCGCTAACTACGTCGAGCAGCGCGGTAAATACGTCATGGTCAAGCTTGT 509  
DB 481 AGGAAGCCAGCTAACTACGTCGAGCAGCGCGGTAAATACGTCATGGTCAAGCTTGT 540  
  
QY 510 TCGGAATCATTGGCGTAAAGGTCGTAGGCGGATTTGTAAAGTCAGGTTGAAACTGC 569  
DB 541 CCGGAATCATTGGCGTAAAGGTCGTAGGCGGATTTGTAAAGTCAGGTTGAAACTGC 600  
  
QY 570 GGGCTCAACCGCTGGCTGCACTGAACTACAACTGAGTGGAGTTGGAGAGCAAGTGG 629  
DB 601 CGGCTCAACCGCTGGCTGCACTGAACTGAGTGGAGTTGGAGAGCAAGTGG 660  
  
QY 630 AATCCAGGTGTAGCGTGAATCGTAGATATCTGGAGGAACCACTGTCGAGGCGGA 689  
DB 661 AATCCAGGTGTAGCGTGAATCGTAGATATCTGGAGGAACCACTGTCGAGGCGGA 720  
  
QY 690 CTGCTGGCTCAACCTGAGGCAAGCGTGGTAGTAACCGGATTAGATA 749  
DB 749 CTGCTGGCTCAACCTGAGGCAAGCGTGGTAGTAACCGGATTAGATA

DB 721 CTCCTGGCTGTAACTGACGCTGAGGCGCAAGGCTGGGAGCGAAGCAGGATTAGATA 780  
QY 750 CCCCGTAATCCACCCCTAAACGTTGTCTACCAAGTTGTTGGGGTTT--AACCTCAG 807  
DB 781 CCTGGTGTCCACCGCTAAACGTTGTCTACCAAGTTGTTGGGGTTT--AACCTCAG 840  
QY 808 TAACGAACCTAACGTTAAAGTAGACCCCTGGGAGCTATGCTCGAAGAGTGAACCTCA 867  
DB 841 TGCTCAGCAACGCTAAAGCACTCCCGCTGGGAGTACGCGCCGCAAGGCTGAACCTCA 900  
QY 868 AAGGAATTGACGGGGTCCGCAACAGCGTGGAGCATGTGTTAATTCGATGATACCCC 927  
DB 901 AAGGAATTGACGGGGTCCGCAACAGCGTGGAGCATGTGTTAATTCGAGCAACGCG 960  
QY 928 AAAAACTCACTGGCTGACATGATGATGAATC-----ATGTAGATATATAGGCT 982  
DB 961 AAGAACCTTACAGGCTCTGACATCCCTGACAAACCTTAGAGATAGGCGTCTCCCTTC 1020  
QY 983 TCGGCAAGATTACAGGCTGCTGATGTTGCTGTCAGCTGTCGTGAGATGTTGGTT 1042  
DB 1021 GGGGACAGGTTGACAGGTTGTCATGTTGTCGTCAGCTGTCGTGAGATGTTGGTT 1080  
QY 1043 AAGTCCCGCAACGAGCGCAACCCCTATGTCATGTTGCTA-CCTAAGTTGGGCACTGTA 1101  
DB 1081 AAGTCCCGCAACGAGCGCAACCCCTGACCTAGTTGTCAGCATTCAGTTGGGCACTGTA 1140  
QY 1102 CGAACTCCCGTGAACAAACCGGAGGAGGCGGATGACGTCGCAATCCTCATGCTT 1161  
DB 1141 GGTGACTCCGCTAAAGTCGAGGAGGTTGGGATGACGTCGCAATCCTCATGCTT 1200  
QY 1162 ATGTCCAGGCGCAACGCTGCTCAATGCGCATACAGAGGTCGCAACCTCGCAAG 1221  
DB 1201 ATGACTGGGTACACGCTGCTCAATGGTGGTACAAAGGCGAGCGAAGCGGAGGT 1260  
QY 1222 GAGCTAATCTTAAAGTGGTCCAGTTCGAGTTGGGTCGCAACCTCGACCCATGA 1281  
DB 1261 GAGCGAATCCCAAAACCACTCTCAGTTGAGTTCGAGGTCGCAACCTCGCATGA 1320  
QY 1282 AGTCGGAATCCTAGTAACTCGGATCAGCATGCGCGGTGAATACGTTCCCGGACCTT 1341  
DB 1321 AGCGGAATCCTAGTAACTCGGATCAGCATGCGCGGTGAATACGTTCCCGGACCTT 1380  
QY 1342 TACACCGCGCTGCAACCACTGAGTGGGAGCAGCCGAGTGGTCTTGTAAACCT 1401  
DB 1381 TACACCGCGCTGCAACCACTGAGTGGTGAACCCGAGTGGTGAACCCGAGTGGT 1440  
QY 1402 A-AGGAGCACTACTAAGGTGAACCTCGTAAAGGGGTGAAGTCGTAACAGGTACC 1459  
DB 1441 ACGGAGCCAGCCGCAAGGTGGGCAAAATGATGGGTTGAAGTCGTAACAGGTAGC 1499

RESULT 8  
US-08-114-695A-1  
; Sequence 1, Application US/08114695A  
; Patent No. 5508193  
; GENERAL INFORMATION:  
; APPLICANT: Mandelbaum, Raphael T.  
; APPLICANT: Wackett, Lawrence P.  
; TITLE OF INVENTION: DEGRADATION OF S-TRIAZINES IN SOIL AND  
; TITLE OF INVENTION: WATER  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCHWEGMAN, LUNDBERG & WOESSNER, P.A.  
; STREET: 3500 IDS CENTER  
; CITY: MINNEAPOLIS  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

RESULT          9
US-08-757-653-158
? Sequence 158, Application US/08757653
? Patent No. 5843669
? GENERAL INFORMATION:
? APPLICANT: Kaiser, Michael W.
? APPLICANT: Lyamichev, Victor I.
? APPLICANT: Lyamichev, Natasha
? TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
? TITLE OF INVENTION: Thermostable FEN-1 Endonucleases
? NUMBER OF SEQUENCES: 190
? CORRESPONDENCE ADDRESS:
? ADDRESSES: Medlen & Carroll, LLP
? STREET: 220 Montgomery Street, Suite 2200
? CITY: San Francisco
? STATE: California
? COUNTRY: United States Of America
? ZIP: 94104
?

```











US-08-632-470-53  
; Sequence 53, Application US/08632470  
; Patent No. 5976791  
; GENERAL INFORMATION:  
; APPLICANT: RAULAT, CLAUDE  
; APPLICANT: RAULAT, DIDIER  
; TITLE OF INVENTION: NUCLEOTIDE FRAGMENTS CAPABLE OF  
; TITLE OF INVENTION: HYBRIDIZING SPECIFICALLY TO RICKETTSIA RDNA OR RRNA AND  
; TITLE OF INVENTION: THEIR USE AS PROBES OR PRIMERS  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OLIFF & BERRIDGE  
; STREET: P.O. BOX 19928  
; CITY: ALEXANDRIA  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/632,470  
; FILING DATE: 08-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERRIDGE, WILLIAM P  
; REGISTRATION NUMBER: 30,024  
; REFERENCE/DOCKET NUMBER: WPB 38238  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)-836-6400  
; TELEFAX: (703)-836-2787  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1484 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-632-470-53

Query Match 52.3%; Score 772.6; DB 2; Length 1484;  
Best Local Similarity 74.1%; Pred. NO. 5.2e-269;  
Matches 1082; Conservative 0; Mismatches 337; Indels 42; Gaps 7;  
QY 1 gatcatgctcagaactaacgtgctgctgcttcttaaacatgcaagtcgagcggttag 60  
DB 15 gatctggctcagattgaaacgctagcggtatccttaaacatgcaagtcgagcggttag 74  
QY 61 caatacct-----agcggcgaacgggtgagtaacacgtggttaactctt 102  
DB 75 CAGGGAGNCTTCTCTGCGCGCGAGAGTGGCGGAGTGAATGCGTAGGAATCTA 134  
QY 103 cctccgagctctgggataactttccgaaggaagtaataaccggatagctcctgttgatc 162  
DB 135 CTTGTAGTGGGGGATAAACCCTGGGGAACCTCGGGTAAATACCGTAAATCTCTTTGGAGC 194  
QY 163 acaagattgatagtgaaagattatgcttgagatagcccgccgagtagtagt 222  
DB 195 AAAGCGGGGGAT--CTTCGGACCTCGTGTATTAAGATGAGCCTACGTGCGATTAGCTTGT 252  
QY 223 tggtaggtaattggctcaccaggagacgatacggtagccggcctgagaggggtgctccggcc 282  
DB 253 TGGTGGGGTAATGGCTTACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGACGATCAGCC 312  
QY 283 acaatgaactagacacagtgctcactcctcagggagcagcagtaagaatcttgctc 342  
DB 313 ACATGGGATGAGACNCGGCGCAGACTCCTACGGGAGGACGAGTGGGGAATATTGGAC 372  
QY 343 aatgggggaacccctgaacgagcgcgcgtgaacgaagaggtcttcgattgttaa 402

Db 373 AATGGGGGAAACCCCTGATCCACCAATGCCGCGTGTGTGAAGAAGCCCTTCGGGTGTAAA 432  
QY 403 gtccattaggcaggaataaagcagcaatgtga-----tgatggttacctgc 449  
Db 433 GCACCTTCGGTGGGGAAGAAATCTCAAGGTTAATATCTCTNGGCGGTTGACGTTACCCAC 492  
QY 450 cta--aacgacccgctaactacgtgcccagcgccggtataacgtatggtgaagcgtt 507  
Db 493 AGAAGAAGCACTGGCTAACTCTGTGCCAGCAGCCGCGGTAAATACAGAGAGTCAACGCGTT 552  
QY 508 gttcgaataatcattggcgtaaaagggtcgttagcggaattgtaagtcaggtgtgaaact 567  
Db 553 AATCGGAATCACTGGGCGTAAAGCGCGTGGTGGATATTTAACTCGGATCGAAGGCC 612  
QY 568 gcgggctcaacccgtggcctgcacttgaaactacaagtcctgaggttgggagagcaagt 627  
Db 613 CTGGCTNNNCTGGGAATTGCACCCGATCTCTTCTGAGTATGCTAGAGGAANT 672  
QY 628 ggaattccaggtgttagcgtgtaaatgctagatatacttgaggagaacacagtgagcaagc 687  
Db 673 GGAATTCGGGTGTAGCGGTGAATGCGTAGATATCGGAAGAAGCAACCAAGTGGCGAAGC 732  
QY 688 gacttgcctggtctcaaaactgacgtgagcgacgaagcgtggtgagtaaacgggattaga 747  
Db 733 GACTTCCTGGACCAATACTGACACTGAGCGCGCAAGCGTGGGGAGCAAAACAGGATTAGA 792  
QY 748 taccgcggttaattccacgcccctaaacgttgtctaccagttgttgggggttttaacccctcag 807  
Db 793 GACCTGTGTAGTCCACGCGNGTCAACGATGAGAACTAGCTGTGGGAAGTTCCCTCTTATG 852  
QY 808 taacgaactaacggatgaagtagacccgctgagactatgctcgaagagagaactca 867  
Db 853 TAGCGAAGCTAACCGTTAAAGTTCTCCGCTGGGAGTACGGCCGCAAGTGTAAACATCA 912  
QY 868 aaggaaattgacgggggtccgcacaagcgtgagcgatgtggttattcgtatgataccccc 927  
Db 913 AAGAAATTGACGGGGCGCGCACAAAGCGGTGGAGCATGTGGTTTAACTCGATGCAACGCG 972  
QY 928 aaaaacctcaacctgggcttgacatggaatcgtgaatcatgtagagatatagacccctcgg 987  
Db 973 AAAAAACCTTACCTACCTTGGATCTCTCGGAACCTGTGAGAGATGATTTGGTGCCTTCGG 1032  
QY 988 ----cagattcacaggtgctgcaatgtgttcgctcagctcgtgctgagatgttgggttaa 1044  
Db 1033 GAACCGAGTGCAGAGTGTGTCATGCGTGTGTCAGTGTGTCGAGATGTGGGTAA 1092  
QY 1045 gtcgcgcaacgagcgcaacccctctatcgtatgttctacc---ttaagttgggcactggtta 1101  
Db 1093 GTCCCGTAACGAGCGCAACCCCTCGTCTTAGTTGCCAGCGAGTCAAGTCGGGAACCTTAA 1152  
QY 1102 cgaactgcccgtgcagaaacccgaggaagcggtgagatgacgtcaaatcctcatggtccttt 1161  
Db 1153 GGAGACTGCGGTGTATAAACCGGAGGAGTGGGATGATGTCAAGTCATCATGTCGCCCTT 1212  
QY 1162 atgtccagggccacacagctgctacaatggcgatcacagaggtcgcccaactcgcaagag 1221  
Db 1213 AGGGTAGGGCTACACAGCTGTACAAATGGGAGTACAAGGGGTGCCAAGCGCGAGGT 1272  
QY 1222 ggaacttaattctctaaaggctcgggtcccgatggtgggtctgcaactcgacccactga 1281  
Db 1273 GGAGCTAATCCCAAGAAACACTGCTGTAGTCCGGATTGGAGTGTGCAACTCGACTCCATGA 1332  
QY 1282 agtcggaatcgctagtaatcgcggatcagatgcgcggtgtaatacgttcccggaacttg 1341  
Db 1333 AGTTGGAATCGTGTAAATCGGGAATCAGCATGTCTCGGTTGAATACGTTCTTCGGGCTTG 1392  
QY 1342 tacacaccccgcttcacacacactgagtggtggggagaccccggaagtggtcttcttgaacct 1401  
Db 1393 TACACACCCCGCTCACACCATGGGAGTGAATGTGTACCAAGAGCGG-GTAGGGCTAACCTT 1451  
QY 1402 aaggagacagactactaaggt 1422  
Db 1452 CGGGAGCGCGCTCACACACCGT 1472



```

Qy 1407 gacagactactaaggtgaactcgtataagggggtgaagtcgtataacaaggtaccgtaaatc 1466
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1452 GGACGGUUAACCGAGGAGUGAUCGAUGNNNNNNNNNNNNNNNNNNNGUAAACAAGNNNNNNNN 1511
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1467 ga 1468
      | |
Db 1512 GA 1513

      RESULT 14
      US-08-299-810A-27
      ; Sequence 27, Application US/08299810A
      ; Patent No. 5721097
      ; GENERAL INFORMATION:
      ; APPLICANT: Rossau, Rudi
      ; APPLICANT: Van Heuverswyn, Hugo
      ; TITLE OF INVENTION: HYBRIDIZATION PROBES FOR THE
      ; TITLE OF INVENTION: DETECTION OF BRANHAMELLA CATARRHALIS STRAINS
      ; NUMBER OF SEQUENCES: 28
      ; CORRESPONDENCE ADDRESS:
      ; ADDRESSEE: Merchant & Gould
      ; STREET: 3100 No. 5721097west Center
      ; CITY: Minneapolis
      ; STATE: MN
      ; COUNTRY: USA
      ; ZIP: 55402
      ; COMPUTER READABLE FORM:
      ; MEDIUM TYPE: Floppy disk
      ; COMPUTER: IBM PC compatible
      ; OPERATING SYSTEM: PC-DOS/MS-DOS
      ; SOFTWARE: PatentIn Release #1.0, Version #1.25
      ; CURRENT APPLICATION DATA:
      ; APPLICATION NUMBER: US/08/299,810A
      ; FILING DATE: 01-SEP-1994
      ; CLASSIFICATION: 435
      ; ATTORNEY/AGENT INFORMATION:
      ; NAME: Hillson, Randall A.
      ; REGISTRATION NUMBER: 31,838
      ; REFERENCE/DOCKET NUMBER: 8076.70-US-WO
      ; TELECOMMUNICATION INFORMATION:
      ; TELEPHONE: 612-332-5300
      ; TELEFAX: 612-332-9081
      ; INFORMATION FOR SEQ ID NO: 27:
      ; SEQUENCE CHARACTERISTICS:
      ; LENGTH: 1485 base pairs
      ; TYPE: nucleic acid
      ; STRANDEDNESS: single
      ; TOPOLOGY: linear
      ; MOLECULE TYPE: DNA (genomic)
      ; ORIGINAL SOURCE:
      ; ORGANISM: Branhamella catarrhalis
      ; IMMEDIATE SOURCE:
      ; CLONE: 16S rRNA Gene
      ; US-08-299-810A-27

```

	Query Match	52.0%	Score 768.4;	DB 1;	Length 1485;
	Best Local Similarity	75.1%;	Pred. No. 1.7e-267;		
	Matches 1075;	Conservative	0; Mismatches 322;	Indels 34; Gaps	8;
Qy	51	agcgggtagcaatcacctagcgacgcgaacgggtgagtacaacgcttggtaatcttccctcagag	110		
Db	35	AGCTTGCTTCGTACTTTAGTGGCGGACGGGTGAGTAATGCTTAGGAANTCGCCTAGTAG	94		
Qy	111	tctgggataacttcgccaaaggaaagcttaacaccggaatgacctgttgtgatccacaagatt	170		
Db	95	TGGGGGATACTTGGGGAAACCCCAAGCTAATACCGCATACGACCTACGGGTCAAGG--	151		
Qy	171	tgtaggttaaagatttatgtcttggaatgagcccgcgccgatttagctaagttgggtgagg	230		
Db	152	----GGGCTTTTACGCTCTCGCTATTATGATAGACCTAAGTCGAGTAGTGGTTGGTGGGG	207		

[illegible]

Qy	170	ttgataagtaag-----atttattgcttgagatgagccgcgcccatttagcta	220
Db	188	ACTTTTACCAAGCGCGCTTTTACGTCGTTACTTTGGATGGGCTCGCTCATTTAGCTA	247
Qy	221	qttaggtca-ogtaatggctcaccaaagcaacgatcgttagcggcctgagaggggtctcg	279
Db	248	GTGGTGACCGGTAAATGGCGCACCACAGCAACGATCATGTACCGCGTCTGAGAGGATGGAGCG	307
Qy	280	gccacaatggaactgagacacggtcccaactctctcagcgagggcagcagttaagaatacttgg	339

pt 399

[illegible]

		108	CTCTTCGGAGGACAAAGAAAAACAGGTGGTGTGATGGCTGTGCTAGCTCGTGTCTGTGAGATG	108
Db		1026		
Qy		1036	tgttggttaagtccgcgaacagacgcaacccctatctatgttc---taccttaagtgttg	109
Db		1086	TGTGGTTAAAGTCCCGCAACGAGCGCAACCCCTATCATTTGTGCCAGCAGTAAACGGTGG	114
Qy		1093	gcacttggtagaagaactgccggtgacaaaccggagggaagcggggtagcgtcaaatccctc	115
Db		1146	GAACTCAAAATGAGACTCCCGCGGACACGCGGAGGAGGCGGGGATGACGTCAAGTCATC	120
Qy		1153	atgacgtttaaagtccaggggccacacacagctgtctacaatggccgatacacagagggtcgccaac	121
Db		1206	ATGCCCTTACGTCTCTGGGTACACAGCTACTCAATGGGATACACAGAGGGGAAGCAAG	126

Db	1326	GAGCGATCGGACCGAACCCAAAAAATATCCCACAGTTCGGATTTCAGAGCTGCAACTCG	1325
Qy	1273	accccatgaagtcggaatcgctagtaatcgcgatacagcatgcgcggtggaatcaccttcc	1332
Db	1326	CCTCATGAAGTCGGAAATCGCTAGTAAATCGGAGGTCAGCATACTCGCGTCAATACGTTCC	1385
Qy	1333	cggacactgtacacacccgcctcacaccactgattggggagcaccgccgaagtggctttt	1392
Db	1386	CGGCGCTTGTACACACCGCCGCTCACACCACGAAAGTCATTTCACACCCGAAGCGGCTAA	1445
Qy	1393	gttaacgctaagagagacagactactaaagtgaaactogtaagggggtgaagtcgtaaca	1452
Db	1446	G--GGCTATTGGAACCGACCCCTCTAAGGTGGGGCGGATGATTGGGTGAAGTCGTAACA	1503
Qy	1453	aggtacc	1459
Db	1504	AGGTAGC	1510

Search completed: March 27, 2001, 08:19:18  
Job time: 5145 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 06:21:57 ; Search time 2517.78 Seconds  
(without alignments)  
4110.788 Million cell updates/sec

Title: US-09-380-826A-1  
Perfect score: 1477  
Sequence: 1 gatcaggctcgaactaac.....ccgtaaatcgattcctgcag 1477.

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues 15983484

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estrol:\*  
75: em\_estrol2:\*  
76: em\_estrol3:\*  
77: em\_estrol4:\*  
78: em\_estrol5:\*  
79: em\_estrol6:\*  
80: em\_estrol7:\*  
81: em\_estrol8:\*  
82: em\_estrol9:\*  
83: em\_estrol10:\*  
84: em\_estrol11:\*  
85: em\_estrol12:\*  
86: em\_estrol13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estpl6:\*  
 118: em\_estpl7:\*  
 119: em\_estpl8:\*  
 120: em\_estrol14:\*  
 121: em\_estrol15:\*  
 122: em\_estrol16:\*  
 123: em\_estrol17:\*  
 124: em\_estrol18:\*  
 125: em\_estrol19:\*  
 126: gb\_est58:\*  
 127: gb\_est59:\*  
 128: gb\_est60:\*  
 129: gb\_est61:\*  
 130: gb\_est62:\*  
 131: gb\_est63:\*  
 132: gb\_est64:\*  
 133: gb\_est65:\*  
 134: gb\_est66:\*  
 135: gb\_est75:\*  
 136: gb\_est76:\*  
 137: gb\_est77:\*  
 138: gb\_est78:\*  
 139: gb\_est79:\*  
 140: gb\_est80:\*  
 141: gb\_est81:\*  
 142: gb\_est82:\*  
 143: gb\_est83:\*  
 144: gb\_est84:\*  
 145: gb\_est85:\*  
 146: gb\_est86:\*  
 147: gb\_est87:\*  
 148: gb\_gss1:\*  
 149: gb\_gss2:\*  
 150: gb\_gss3:\*  
 151: gb\_gss4:\*  
 152: em\_gss1:\*  
 153: em\_gss2:\*  
 154: em\_gss3:\*  
 155: em\_gss4:\*  
 156: gb\_gss5:\*  
 157: gb\_gss6:\*  
 158: gb\_gss7:\*  
 159: gb\_gss8:\*  
 160: gb\_gss9:\*  
 161: gb\_gss10:\*  
 162: gb\_gss11:\*  
 163: gb\_gss12:\*  
 164: gb\_gss13:\*  
 165: gb\_gss14:\*  
 166: gb\_gss15:\*  
 167: gb\_gss16:\*  
 168: gb\_gss17:\*  
 169: gb\_gss18:\*  
 170: gb\_gss19:\*  
 171: gb\_gss20:\*  
 172: gb\_gss21:\*  
 173: gb\_gss22:\*  
 174: gb\_gss23:\*  
 175: gb\_gss24:\*  
 176: em\_gss5:\*  
 177: em\_gss6:\*  
 178: em\_gss7:\*  
 179: em\_gss8:\*  
 180: em\_gss9:\*  
 181: em\_gss10:\*  
 182: em\_gss11:\*  
 183: em\_gss12:\*  
 184: em\_gss13:\*  
 185: em\_gss14:\*  
 186: em\_gss15:\*  
 187: em\_gss16:\*  
 188: em\_gss17:\*  
 189: em\_gss18:\*

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	436.8	29.6	800	165	AQ960400 LERFD51TR
C 2	424	28.7	716	165	AQ957363 LERAP36TR
C 3	407.4	27.6	720	190	AL093421 Arabidops
C 4	389	26.3	746	165	AQ957362 LERAP36TF
C 5	383.6	26.0	741	165	AQ957181 LERAO33TR
C 6	378.8	25.6	757	165	AQ957182 LERAO34TR
C 7	376.6	25.5	676	175	B73824 T27K4TR TAM
C 8	360.4	24.4	693	165	AQ956298 LERAJ19TR
C 9	359.6	24.3	676	165	AQ969911 LERJR28TR
C 10	356.6	24.1	659	165	AQ969996 LERJR90TF
C 11	352.4	23.9	677	87	AW221851 EST298662
C 12	350.6	23.7	616	190	AG019017 Homo sapi
C 13	346.8	23.5	675	165	AQ956297 LERAJ19TF
C 14	344	23.3	659	95	AQ956297 LERAJ19TF
C 15	343	23.2	603	37	AW833978 QVO-TT000
C 16	340.4	23.0	768	165	AV531914 AV531914
C 17	340	23.0	637	165	AQ957272 LERAO84TR
C 18	339.6	23.0	687	165	AQ962145 LERGD59TR
C 19	339	23.0	632	165	AQ955848 LERAG49TR
C 20	337.4	22.8	618	165	AQ967817 LERIV15TF
C 21	336.6	22.8	601	174	AQ969912 LERJR28TR
C 22	334.2	22.6	660	92	B27166 T2C12TR TAM
C 23	333.6	22.6	592	190	AW600903 3B5 CDNA
C 24	332.4	22.5	599	165	AG019018 Homo sapi
C 25	332	22.5	648	165	AQ968703 LERJE31TF
C 26	331.6	22.5	561	190	AQ957342 LERAP25TF
C 27	331.6	22.5	616	107	AG019935 Homo sapi
C 28	331.6	22.5	673	148	BE356988 DGL_146_B
C 29	330.8	22.4	604	105	AF075762 AF075762
C 30	327.6	22.2	662	94	BE204135 EST396811
C 31	327.2	22.2	639	165	AW736167 EST332153
C 32	324.4	22.0	610	165	AQ956183 LERAI55TR
C 33	319.8	21.7	553	107	AQ961463 LERPL40TR
C 34	319	21.6	649	93	BE362383 DGL_86_H0
C 35	315.2	21.3	567	165	BE362383 DGL_86_H0
C 36	314.2	21.3	780	190	AW675839 SWYDZ5CAU
C 37	312	21.1	584	165	AQ961462 LERPL40TF
C 38	311.6	21.1	583	165	AL094925 Arabidops
C 39	310.6	21.0	503	92	AQ964113 LERCT38TF
C 40	310	21.0	676	105	AQ967819 LERIV15TR
C 41	308.2	20.9	766	111	AW563518 LGL_239_H
C 42	307.4	20.8	714	105	BE187622 EST336183
C 43	304.4	20.6	540	165	BE661742 969 Gmaxs
C 44	302.4	20.5	899	174	BE202334 SWOVIFCAR
C 45	299.4	20.3	553	165	AQ968704 LERJE31TR
					B10095 F8G13-T7 IG
					AQ966081 LERII07TF

## ALIGNMENTS

RESULT 1  
 AQ960400/c  
 LOCUS  
 DEFINITION LERFD51TR LERA Arabidopsis thaliana genomic clone LERFD51, DNA  
 800 bp  
 DNA  
 CSS  
 28-JAN-2000  
 ACCESSION AQ960400  
 VERSION AQ960400.1  
 GI:6788101  
 KEYWORDS  
 CSS.  
 thale cress.



[illegible]

RESULT	2				
AQ957363/c					
LOCUS			716 bp	DNA	
DEFINITION		AQ957363		GSS	28-JAN-2000
		LERAP36TR	LERA	Arabidopsis thaliana	genomic clone LERAP36, DNA
					sequence.
ACCESSION		AQ957363			
VERSION		AQ957363.1	GI:6785064		
KEYWORDS		GSS.			
SOURCE					thale cress.
ORGANISM					Arabidopsis thaliana

REFERENCE	1 (bases 1 to 716)			
AUTHORS	Buell, C.R., Lin, X., Pal, G., Barnstead, M., Bowman, C., Utterbach, T., Feldblum, T., Liang, F., Creasy, T. and Fraser, C.M.			
TITLE	Genomic survey sequencing of <i>Landsberg erecta</i> ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: at@tigr.org For additional information, see <a href="http://www.tigr.org/tdb/at/at.html">http://www.tigr.org/tdb/at/at.html</a> Seq primer: TR Class: shotgun.			
FEATURES	source	Location/Qualifiers		
		1..716		
		/organism="Arabidopsis thaliana"		
		/strain="Landsberg erecta"		
		/db_xref="taxon:3702"		
		/clone="LERAP36"		
		/clone_lib="LERA"		
		/note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."		
BASE COUNT	144	a	213	c 154 g 205 t
ORIGIN				
	Query Match	28.7%	Score 424;	DB 165; Length 716;
	Best Local Similarity	77.6%	Pred. No. 1.3e-123;	
	Matches 545;	Conservative	0;	Mismatches 140; Indels 17; Gaps 2;
Qy	275	gtccggcccaatggaactgagacacggtccatactctctacggagcagcagcttaagaa	334	
Db	716	GACCAGCCATGCGAACTAAGACACGCTCCAGACTCTCAGCGGAGGACGATGGGAA	657	
Ov	335	tcttactcaatgggggaacccctgaacgacgcaccccgtaacagagaagactttcga	394	

```
Db 656 TTTTGGCAATGGGCGCAAGCCTGATCCAGCCATTCGCGGTGAGTGAAGAAGGCTTCGG 597
Qy 395 attgaaagtccattagcagcaaaaaaagcagcaaatgtagatgtagtacctccta-- 452
Db 596 GTTGTAAAGCTCTTTCGCAAGGGGAAGAAACTTACTTCTTAATAAAGAGTAGGCTGACG 537
Qy 453 -----aagcaccggctaaactacgtgcagcagcgcggtgtaatacgtatggtg 499
Db 536 GTACCTTGATAGAAGCAGCAGCTTAACCTACGTGCAGCAGCGCGGTAAATAGTAGGGTG 477
Qy 500 caagcgttgcgaatacatctggcgtaaaagggcgtagcgaggtattgtaagtcaggtg 559
Db 476 CAAGCGTTAAATCGGAATTAACGGCGTAAAGCGTTCGCGAGCGGTTGGCAAGTCAGATG 417
Qy 560 tgaacctgcgggtcgaacctggcctgcacttgaactcaaatcaagctggagtttggag 619
Db 416 TGAATCCCGGAGCTCACTTGGGNACTGCGTTCGTAACACTGCCAGACTAGATATGTCAG 357
Qy 620 aggaagtggattccaggttagcgtggaatcgtagatatctggaggaacccagtg 679
Db 356 AGGGGGGTAGAAATTCACGCTGAGCAGTGAATGCGTAGAGATGTGGAGGAATACCAATG 297
Qy 680 gcgaagcgacttgctgctcaaaactgaactgaactgaagcagcaaaagcgtgtagtaaacg 739
Db 296 GCGAAGGCGAGCCCCCTGGGATAATATTGACGCTCATGCACGAAAGCGTGGGAGCAACA 237
Qy 740 ggattagatacccggttaatacagccctaaacgttgctcaccagttgttgggggttta 799
Db 236 GGATTAGATACCTTGGTAGTCCACGCCCTAAACGATGCTACTAGTTGTTGGAGTAA 177
Qy 800 a---ccctcagtaacgaactaacggattaaagtagacacgctgggaactatctcgaaga 857
Db 176 AATCATGAGTAACGACACTAACCGGTGAAGTAGACCGCTGGGAGTAGCGCAAGA 117
Qy 858 gtgaactcaagaagatgacggggtccgcacagcgtgagcatgtgttaattcgt 917
Db 116 TTAAACTCAAAAGGAATTCACGGGGGGCCCCGACACCGCGTGATATGTGATTAATTCG 57
Qy 918 atgatacccccacaaacctcactcgttggcttgacatggatctga 959
Db 56 ATGCAACGCAAAACCTTACCTGGCCCTTGACATGCCACTAA 15

RESULT 3
CNS00MCV 720 bp DNA GSS 28-JUN-1999
LOCUS Arabidopsis thaliana genome survey sequence Sp6 end of BAC T11h14
DEFINITION of TAMU library from strain Columbia of Arabidopsis thaliana,
genomic survey sequence.
ACCESSION AL093421 GI:5294575
VERSION AL093421.1
KEYWORDS GSS.
SOURCE Salanoubat,M., Cholsne,N., Artiguenave,F., Brottier,P., Wincker,P.,
Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 720)
AUTHORS Salanoubat,M., Cholsne,N., Artiguenave,F., Brottier,P., Wincker,P.,
Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 720)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
source Location/Qualifiers
1..720
/organism="Arabidopsis thaliana"
/strain="Columbia"
```

```
/db_xref="taxon:3702"
/clone_lib="TAMU"
/clone="T11h14"
/note="end : Sp6"
BASE COUNT 178 a 175 c 230 g 137 t
ORIGIN

Query Match. 27.6%; Score 407.4; DB 190; Length 720;
Best Local Similarity 77.2%; Pred. No. 2.5e-118;
Matches 547; Conservative 0; Mismatches 156; Indels 6; Gaps 4;

Qy 609 gagttgggagagcaagtggaattccagggtgtagcgggtgaaatgcgtagatatctggag 668
Db 1 GAGTACGGTTGGGCGAGAGGGAATTTCCGGTGGAGCGGTGAATTCGGTAGAGATCGGAAA 60
Qy 669 gaacacagtgccgaagcgacttgcctcaaaactgacgtgagcagcaagaagcgtg 728
Db 61 GAACACCAACCGCGAAGACACTCTGCTGGGCCGACACTGACACTGAGAGAGCAAGAGCTAG 120
Qy 729 ggtagtaaacgggattagatatacccccggtaatccacgcgcctaaacggttctaccagttgt 788
Db 121 GGGAGCGAATGGGATTAGATACCCCACTAGTCTAGCCGTAAACGATGAGTACTAGGCGC 180
Qy 789 tgggggttttaaccc--tcagtaacgaacctaacggattaagtagagacgcgcgggacta 846
Db 181 TGTGCGTATCGACCGCTGCTAGTCTAGCTTAACGCGTTAAAGTATCCGCGCTGGGGAGTA 240
Qy 847 tgcgcgaagtgtaaaactcaaggaattgacgggggtccgcacaaagcgggtggagcatgt 906
Db 241 CTTTCGCAAGAATCAAACTCAAGGAATTTGACGGGGGGCCGCCACAACGGGTGGAGCATGT 300
Qy 907 ggtttaattcgatatacccccaaaacctcactcgggtgggttgacatggtatcgaatcgt 966
Db 301 GGTTAATTTCGATCAAAAGCGAAGAACCTTACCAGGGCTTGACATGCGCGGAATCCTCTT 360
Qy 967 -agagatatagaccttcggggagat--tcacagggtcgtcgtatggttgcgtcagtcgt 1023
Db 361 GAAAGAGAGGGGTGCCCTTCGGGAACCGGACACAGGTGTCATGCGTGTGCTGCTGAGCTCG 420
Qy 1024 tgcgtgagatgttgggttaagtcgccgcacagcgcgaacccctatcgtatgttgcacc 1083
Db 421 TGCGTAAAGTGTGGGTTAAGTCCCGCAACGAGCGCAACCCCTCGTGTAGTTGCCACC 480
Qy 1084 -ttaagttgggacttgggtacgaaactcgcggtgacaaaccggaaggaagcgggagtgacg 1142
Db 481 GTTGAGTTTGGAAACCTGAACAGACTGCCGGTGTAGTAAAGCGGAGGAGGTGAGGATGACG 540
Qy 1143 tcaatcctcgtcgttcttgcaggggccacacacgctgacacagtgccgatacagaag 1202
Db 541 TCAAGTCATCATGCCCCCTTATGCCCCGGGCGACACACGTGTACATGGCGGGGACAAG 600
Qy 1203 ggtcgccaaactcgcaagaggagcttaactctctaaagtcggtcccgatcggatggggt 1262
Db 601 GGTGCGGATCCCGGAGGGTGCAGCTAACTCCAAAAACCCGTTCTCAGTTCCGATTCGAGG 660
Qy 1263 ctgcaactcgccccatgaatcggaatcgtcgtatcgtatcgcggatcagc 1311
Db 661 CTGCAACTCGCTGCATGAAGCGGAATCGTAGTAATATCGCGCGGTGACG 709

RESULT 4
AQ957362 746 bp DNA GSS 28-JAN-2000
LOCUS LERAP36TF LERA Arabidopsis thaliana genomic clone LERAP36, DNA
DEFINITION sequence.
ACCESSION AQ957362
VERSION AQ957362.1 GI:6785063
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids II;
```

	Brassicales; Brassicaceae; Arabidopsis.	
REFERENCE	1 (bases 1 to 746)	
AUTHORS	Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T., Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.	
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atetigr.org For additional information, see http://www.tigr.org/tldb/at/at.html Seq primer: TF Class: shotgun.	
FEATURES	Location/Qualifiers	
source	1..746	
	/organism="Arabidopsis thaliana" /strain="Landsberg erecta" /db_xref="taxon:3702" /clone="LERAP36" /clone_lib="LERA" /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."	
BASE COUNT	208 a 153 c 225 g 159 t	1 others
ORIGIN		
	Query Match 26.3%; Score 389; DB 165; Length 746; Best Local Similarity 73.0%; Pred. No. 1.8e-112; Matches 534; Conservative 0; Mismatches 181; Indels 17; Gaps 2;	
Qy	84 agtaacacggtgaatcttccctcgcagtgctggataacttcgaaaggaaagcaaatcatc 143	
Db	1 AGTAATATATCGGAACGCTGGTGTATGGGGTAACTAGTCGAGAAGATTAGCTAATAAC 60	
Qy	144 cgtagtcctgttgatcacagaatttgtatagtgtaaagtattttgttgagatgagc 203	
Db	61 CGCATAGCCCTCAGGGGAAAAGTAGGGGATC--TTCCGACCTTACGTTATAAGAGCGGC 118	
Qy	204 ccgcgcgcgattagctagttggtgaggttaattggctaccacaagcgacgtcgtgagccgg 263	
Db	119 CGATATCTGATTAGCTAGTTGCTGGGGTAAATGGCTTACCAGGCTTCGATCAGTAGCTGG 178	
Qy	264 cctgagaggggtgccgccaataatgaactgagacacggttcatactctcagggagacca 323	
Db	179 TCTGAGAGGACGACACGCCACTGGAACCTAGACACGGTCCAGACCTCTACGGGAGCCA 238	
Qy	324 gcagttaagaatcttgcctaagtggggaacctgaagcagcagccgcgtgaacgaag 383	
Db	239 GCAGTGGGGAATTTTGACAATGGCGCAAGCCCTGATCCAGCCATTTCCCGCTGAGTGAAG 298	
Qy	384 aagttcttcggattgtaaaattcattagcgcagaaaaataaagcagcaatgtgatgatgt 443	
Db	299 AAGGCTTCGGGTGTAAACCTCTTCCGAAGGGAAGAAACTTACTTTCTAATAAGAG 358	
Qy	444 acctgoccta-----aagcaccggctcaactacgtgcagcagccgggttaa 488	
Db	359 TGAGGCTTGACGGTACCTTGATATAAGAACACCGCTTAACCTGTCGACGAGCGCGGTAA 418	
Qy	489 tacgtatggtcgaagcgttgttcgaatcattggcgttaagggtgcgtagcggatttg 548	
Db	419 TACGTAGGGTGGCAGCGTGAATCGGAATTAATCGGGCTTAACCGGTGCGCAGCGGTTGG 478	
Qy	549 taegtccaggtgtgaaaactcggggctcaaccgcctggcctgcacttgaaactacaagtctg 608	
Db	479 CAAGTCAGATGTAATCCCGAGCTCAACTTGGGAACCTCGCTTTCAAACGCCAGACTA 538	
Qy	609 gagtttgggaggaagtggaattcccaaggtgtagcgggtgaaatgcgtagatatctggag 668	
Db	539 GAATATGTCAGAGGGGGTGAATTCACGCTGTAGCAGTGAATGCGTGAAGATGTGGAG 598	

Qy	669 gaacaccagttggcaagcgactgtcgtggctcaaaaactgacgtcgagggcacgaaacgtg 728	
Db	599 GAATACCAATGGCGAAGCAGCGCCCCCTGNGATAATAATTGACGCTCATGACGAAGAGCGTG 658	
Qy	729 ggtagtaaacgggattagatcacccccggtaatacccacgcctcaaacggttgtctaccagttgt 788	
Db	659 GGGAGCAACAGGATTAGATACCTGGTAGTCCACGCCCTAAAAAGATGTCTACTAGTTGT 718	
Qy	789 tgggggttttaa 800	
Db	719 TGGTGGAGTAAA 730	
RESULT	5	
LOCUS	AQ957181	741 bp DNA GSS 28-JAN-2000
DEFINITION	LERAO33TR LERA Arabidopsis thaliana genomic clone LERO33, DNA sequence.	
ACCESSION	AQ957181	
VERSION	AQ957181.1	GI:6784882
KEYWORDS	GSS	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II Brassicales; Brassicaceae; Arabidopsi.	
AUTHORS	1 (bases 1 to 741) Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T., Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.	
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atetigr.org For additional information, see http://www.tigr.org/tldb/at/at.html Similar to A. thaliana chloroplast sequence (GB:AP000423) Seq primer: TR Class: shotgun.	
FEATURES	Location/Qualifiers	
source	1..741	
	/organism="Arabidopsis thaliana" /strain="Landsberg erecta" /db_xref="taxon:3702" /clone="LERO33" /clone_lib="LERA" /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."	
BASE COUNT	137 a 240 c 169 g 195 t	
ORIGIN		
	Query Match 26.0%; Score 383.6; DB 165; Length 741; Best Local Similarity 74.6%; Pred. No. 9.5e-111; Matches 525; Conservative 0; Mismatches 169; Indels 10; Gaps 3	
Qy	190 cttgtgagatgagccgcgcgctgattagctagttggtgaggtaatgctcaccacgacga 249	
Db	697 CCCCAGGAGGGGTCGCGCTGTAGTAGCTTGGTGAGCAATAGCTTACCANGCGCA 638	
Qy	250 cgatcggtagccgcgcgtgagaggggtgtccggcccaaatggaactgagacacggtccatcac 309	
Db	637 TGATCAGTAGCTGGTCCGAGAGGATGATCACCCACACTGGGACTGAGACACGGCCAGAC 578	
Qy	310 tcctacgggagcgacgttaagaatcttgcataatgggggaaacccctgaagcagcgacg 369	
Db	577 TCCTACGGGAGCGACGTCGGGGAAATTTTCGCAATGGCGAAAGCCTGACGGAGCAANTG 518	



REFERENCE 1 (bases 1 to 676)  
 AUTHORS Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter, J.C.  
 TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3  
 JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: T27K4TF  
 Contact: Steve Rounsley  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: rounsley@tigr.org  
 Seq primer: M13 Reverse  
 Class: BAC ends  
 High quality sequence stop: 676.

FEATURES  
 source  
 1..676  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="T27K4"  
 /clone\_lib="TAMU"  
 /sex="hermaphrodite"  
 /note="vector: BelorACII; Site\_1: HindIII; Site\_2: HindIII  
 ; Produced by Rod Wing"  
 BASE COUNT 170 a 164 c 213 g 128 t 1 others  
 ORIGIN

Query Match 25.5%; Score 376.6; DB 175; Length 676;  
 Best Local Similarity 76.2%; Pred. No. 1.6e-108;  
 Matches 515; Conservative 0; Mismatches 155; Indels 6; Gaps 4;

Qy 610 agtttgagaggaagtggaattccaggtgtagcgtggaatcgtagatatctggagg 669  
 Db 1 AGTACGGTAGGGGACAGAGGGAATTTCCGGTGGAGCGGTGAATCGGTAGAGATCGGAAG 60  
 Qy 670 aacaccagtggcgaaggcgacttgcgtctcaaaactgacgtgagcagcaaaagcgtgg 729  
 Db 61 AACACCAACGGGAAGCACTCTCTGGCCGACACTGACACTGAGAGCAAGAGCTAGG 120  
 Qy 730 gtatgaacgggattagatcccccggttaatccacgcccctaaccgtttgtctaccagttgtt 789  
 Db 121 GGAGCGAATGGATTAGATACCCAGTAGTCTAGCCGTAAACAGATGGATAGTACGGCGCT 180  
 Qy 790 gggggttttaaccc--tcagtaacgaacctaacggttaagtagacgcgcctggggaactat 847  
 Db 181 GTGCGTATCGACCCGTGCGAGTGTGTAGCTAACCGGTTTAAGTATCCCGCCCTGGGGAGTAC 240  
 Qy 848 gctcgcaagagtgaactcaaaaggaattgacggggttcgcgcacaaagcgtgtgagcatgtg 907  
 Db 241 GTTCGCAAGATGAACATCAAGGAATTGACGGGGCCCGCACAAAGCGGTGGAGCATGTG 300  
 Qy 908 gtttaattcgatgatacccccaaaacctcaacctgggcttgacatggatctgaate-atgt 966  
 Db 301 GTTTAAATTTCGATGAAACGCGAAGAACCTTTACAGGGCTTGACATGCGCGGAATCCTCTTG 360  
 Qy 967 agagatatagccttcggg--cagattcacaggtgctgcattggttgcctcaagctcgt 1024  
 Db 361 AAAGAGAGGGGTGCTTCGCGGAACCGGACACAGGTGTGTCATGGCTGCTCGTCAGCTCGT 420  
 Qy 1025 gtcgtgagatgtgggttaagtcccgcaacgagcgcaacccctatcgtatgttgctacc- 1083  
 Db 421 GCCGTAAAGGTGTGGTTAAGTTCGCGCAACCGAGCGCAACCCCTCGTGTATTGTCACCG 480  
 Qy 1084 ttaagttgggcaactggttacgaaactgcggttgacaaacccggaggaagcggtggatgcgt 1143  
 Db 481 TTGAGTTTGGAAACCTGTAACAGACTGCCGGTGATAAAGCCGGGAAGGTGAGGATGACGT 540  
 Qy 1144 caaatccctcatggcctttatgtccaggggccacacacgctgtctacaatggccgatacagagg 1203

Db 541 CAAGTCATCATGCCCTTATGCCCTGNGCGACACACGTCGTACAAATGCCGGGACAAAAG 600  
 Qy 1204 gtgcgcaactcgcaagaggagctaatctctaaaagtcggtcccaagttcgcgattggggtc 1263  
 Db 601 GTCCGGATCCCGGAGGTGAGCTAACTCCAAAACCCGCTCTCAGTTCCGATTGCAGCG 660  
 Qy 1264 tgcaactcgaccccat 1279  
 Db 661 TGCAACTCGCCTGCAT 676

RESULT 8  
 LOCUS AQ956298/8 693 bp DNA GSS 28-JAN-2000  
 DEFINITION LERAJ19TR LERA Arabidopsis thaliana genomic clone LERAJ19, DNA sequence.  
 ACCESSION AQ956298  
 VERSION AQ956298.1 GI:6783999  
 KEYWORDS GSS.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 REFERENCE 1 (bases 1 to 693)  
 AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T., Feldblum,T., Liang,F., Creasy,T. and Fraser,C.M.  
 TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atetigr.org  
 For additional information, see <http://www.tigr.org/tdb/at/at.html>  
 Similar to A. thaliana chloroplast sequence (GB:AP000423)  
 Seq primer: TR  
 Class: shotgun.

FEATURES  
 source  
 1..693  
 /organism="Arabidopsis thaliana"  
 /strain="Landsberg erecta"  
 /db\_xref="taxon:3702"  
 /clone="LERAJ19"  
 /clone\_lib="LERA"  
 /note="Organ: Leaf; Vector: phos1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."  
 BASE COUNT 136 a 221 c 157 g 178 t 1 others  
 ORIGIN

Query Match 24.4%; Score 360.4; DB 165; Length 693;  
 Best Local Similarity 77.2%; Pred. No. 2.2e-103;  
 Matches 477; Conservative 0; Mismatches 136; Indels 5; Gaps 3;

Qy 453 aagcaccgctaactacatcgccagcagcgcggttaacgtatgctgaagcgtgttgcg 512  
 Db 626 AAGCATCGGCTAACTCTGTGCCAGCAGCGCGGTAAACAGAGGATGCAAGGTTATCGG 567  
 Qy 513 gaatcattggcgtaaaaggggtgcgtaggcggtattgttaagtcaggtgtgaactgcggg 572  
 Db 566 GAATGATTGGCGTAAGCGTCTGTAGTGGCTTTTAAAGTCGCCGCTCAAAATCCAGGG 507  
 Qy 573 ctcaaacccgtggcctgcacttgaactacaacgtctggagtttggagagggcaagtggaa 632  
 Db 506 CTCAACCCCTGGACAGCGCGGTGGAACACTACCAAGCTTGACGTACGGTAGGGCAGAGGAAT 447  
 Qy 633 tccaggttagcgggtgaactcgtagatatctcgaggacaccacacagtcggaagcgacct 652

Db 446 TTCCGTTGGAGCGGTGAATGCTAGAGATCGAAAGAACACCAACGCGGGAAGCACTCT 387  
Qy 693 gctggtcctaaactcagctgagcgcaagcgtggttagtaaaacggtgattagatcccc 752  
Db 386 GCTGGGCGGACACTGACACTGACAGACGAAGCTAGGGGAGCGAATGGGATTAGATACC 327  
Qy 753 cggtaataccacgcccctaaacgtttgtctaccagttgttgggggttttaacccc--tcagtaa 810  
Db 326 CAGTAGTCTTACCGTAAACCATGGATAGTAGCGCTGTGCTATCGACCCCTGCGATGTC 267  
Qy 811 cgaactaaacggtattagtagacgctggtgactatctcgaagagtgaaactcaag 870  
Db 266 TGTAGCTTAACCGCTTAAGTATCCCGCTGGGAGTACGTTCCGCAAGAAATGAACCTCAAG 207  
Qy 871 gaattgacgggggtccgcacaaacggtggagcgtgtggttaatttcgatatgataccccc 930  
Db 206 GAATTGACGGGGCCCGCACAAAGCGGTGGAGCATGTGTTTAAATTCGATGCAAGCGAAG 147  
Qy 931 aacctcactgggttgacatggtatcgaatc--atgtagagatatagccttcggg-- 987  
Db 146 AACCTTACAGGGCTTGACATGCGCGCAATCCCTCTTGAAGAGAGAGGGGTCCTTCGGGA 87  
Qy 988 cagattcacagtgctgcatgttctgctcagctcgtgctgagatgttgggttaagtc 1047  
Db 86 CCGGACACAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27  
Qy 1048 ccgcacagcgcaacc 1065  
Db 26 CCGCAACGAGCGCAACCC 9

RESULT 9  
AQ969911/c 676 bp DNA GSS 28-JAN-2000  
LOCUS LERJR28TF LERG Arabidopsis thaliana genomic clone LERJR28, DNA  
DEFINITION sequence.  
ACCESSION AQ969911  
VERSION AQ969911.1 GI:6797612  
KEYWORDS GSS.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 676)  
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,  
Feildblum,T., Liang,F., Creasy,T. and Fraser,C.M.  
TITLE Genomic survey sequencing of Landsberg erecta ecotype of  
Arabidopsis thaliana and identification of sequence-based  
polymorphisms  
JOURNAL Unpublished (2000)  
COMMENT Contact: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: atetigr.org  
For additional information, see <http://www.tigr.org/tdb/at/at.html>  
Similar to A. thaliana chloroplast sequence (GB:AF000423)  
Seq primer: TF  
Class: shotgun.  
Location/Qualifiers  
FEATURES  
source 1. .676  
/organism="Arabidopsis thaliana"  
/strain="Landsberg erecta"  
/db\_xref="taxon:3702"  
/clone="LERJR28"  
/clone\_1lib="LERG"  
/note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was  
sheared to 0.4-0.7 kbp before ligation."  
BASE COUNT 133 a 207 c 177 g 159 t  
ORIGIN

Query Match 24.3%; Score 359.6; DB 165; Length 676;  
Best Local Similarity 76.7%; Pred. No. 4e-103;  
Matches 517; Conservative 0; Mismatches 149; Indels 8; Gaps 6;

Qy 755 gtaattccacgcccctaaacgtttgtctaccagttgttgggggttttaacccc--tcagtaa 812  
Db 673 GTAGTCTTACCGTAAACCATGGATAGTAGCGCTGTGCTATCGACCCCTGCGATGCTG 614  
Qy 813 aacctaaacggtattagtagacgctggtgactatctcgaagagtgaaactcaagga 872  
Db 613 TAGTAAACCGCTTAAGTATCCCGCTGGGAGTACGTTCCGCAAGAAATGAACCTCAAGGA 554  
Qy 873 attgacgggggtccgcacaaacggtggagcgtgtggttaatttcgatatgataccccc 932  
Db 553 ATTGACGGGGCCCGCACAAAGCGGTGGAGCATGTGTTTAAATTCGATGCAAGCGAAGAA 494  
Qy 933 cctcactgggttgacatggtatcgaatc--atgtagagatatagccttcggg--ca 989  
Db 493 CCTTACCGAGGCTTGACATGCGCGCAATCCCTCTTGAAGAGAGGGGTCCTTCGGGAACG 434  
Qy 990 gattcacaggtgctgcatgttctgctcagctcgtgctgagatgttgggttaagtc 1049  
Db 433 CGGACACAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 374  
Qy 1050 gcaacgacgcaaccctatcgtatgttgcac--ttaagttggcactggtacgaaact 1108  
Db 373 GCAACGAGCGCAACCCCTCGTGTGTTAGTTGCCACCCGTTGAGTTGGAACCCCTGAACAGACT 314  
Qy 1109 gccggtgacaaacgaggaagcggggtgagtcacgtcaaatcctcctatggtttatgtcca 1168  
Db 313 GCCGCTGATTAAGCCGAGGAGGTGAGGATGACGTCAAGTCATCATGCCCCCTTATGCCCT 254  
Qy 1169 gggccacacacgtgtctacaaatggccgatacacagaggggtgcacactcgaagaggagcta 1228  
Db 253 GGGCGACACACGTGCTTACAAATGGCGGGACAAAGGGTCCGATCCCGGAGGGTGAAGCTA 194  
Qy 1229 atctctaaaagtcggtcccgatggtggtggtggtggtggtggtggtggtggtggtggtggt 1288  
Db 193 ACTCCAAAACCCGCTCCTCAGTTCGGATTGCGAGGTGCAACTCGCTGCTGCATCAAGCCGA 134  
Qy 1289 atcgtagtaatacgcggatcag--catgccgcgggtgtaatacgttcccgacctgtacaca 1347  
Db 133 ATCGTAGTAAATCCCGGTGAGCATACGCGGTGAATTCGTTCCCGGCGCTGTGATACACA 74  
Qy 1348 ccgcccgtcacaccactgagtgaggagcaccgcgaagtggttcttgaaccgttaaggag 1407  
Db 73 CCGCCGCTCACACTATGGGAGTGGCCATCCCGAAGTCGT-TACCTTAACCGCAAGGAG 15  
Qy 1408 acagactactaaag 1421  
Db 14 GGGATGCCGAAG 1

RESULT 10  
AQ969996/c 659 bp DNA GSS 28-JAN-2000  
LOCUS LERJR90TF LERG Arabidopsis thaliana genomic clone LERJR90, DNA  
DEFINITION LERJR90TF LERG Arabidopsis thaliana genomic clone LERJR90, DNA  
sequence.  
ACCESSION AQ969996  
VERSION AQ969996.1 GI:6797697  
KEYWORDS GSS.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 659)  
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,  
Feildblum,T., Liang,F., Creasy,T. and Fraser,C.M.  
TITLE Genomic survey sequencing of Landsberg erecta ecotype of  
Arabidopsis thaliana and identification of sequence-based  
polymorphisms

Unpublished (2000)  
 Contact: Xiaoyang Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: at@tigr.org  
 For additional information, see <http://www.tigr.org/tdb/at/at.html>  
 Similar to A. thaliana chloroplast sequence (GB:AP000423)  
 Seq primer: TF  
 Class: shotgun.

FEATURES  
 source 1..659  
 /organism="Arabidopsis thaliana"  
 /strain="Landsberg erecta"  
 /db\_xref="taxon:3702"  
 /clone="LERJ90"  
 /clone\_lib="LERG"  
 /note="Organ: Leaf; Vector: pUC19JK; Total/genomic DNA was sheared to 0.4-0.7 kbp before ligation."  
 BASE COUNT 130 a 206 c 163 g 160 t  
 ORIGIN

Query Match 24.1%; Score 356.6; DB 165; Length 659;  
 Best Local Similarity 77.3%; Pred. No. 3.6e-102;  
 Matches 497; Conservative 0; Mismatches 139; Indels 7; Gaps 5;

Qy 592 tgcctgctcaaaactgacgtgagcgacgaagcgctgggtagttaaacyggattagataacc 751  
 Db 643 TGCCTGGCCGACACTGACACTGAGAGACGAAGTAGGGGACGAATGGCAATTAGATACC 584  
 Qy 752 ccggtaatccacgcctcaaacctgtctaccagttgttgggggttttaacc--tcagta 809  
 Db 583 CCAGTAGTCTAGCCGTAAACGATGGATACTAGCGCGTGTGCGTATCGACCCGTGCAGTG 524  
 Qy 810 acgaacctaaacgattaaagtacgcgcctggggactatgctcgcaagagtgaactcaaa 869  
 Db 523 CTGTAGCTAACCGCTTAAGTATCCCGCTGGGAGTAGCTTCGCAAGAAATGAAACTCAAA 464  
 Qy 870 ggaattacgggggtccgcacaaacggtgagcagatggtttaattcagatataccccc 929  
 Db 463 GGAATTGACGGGGGCCCGCACACGGGTGGAGCATGTGTTTAATTCGATGCAAGCGAA 404  
 Qy 930 aaacctcacctggcttgacatggaatctgaatc-atgtagagatatagcccttcggg- 987  
 Db 403 GAACCTTACCAGGCGCTTGACATGCGCGGAATCCTCTTGAAGAGAGGGGTGCCCTTCGGGA 344  
 Qy 988 -cagattcaaggtgctgcatggtttgttcagctcgtgtcgtgagatgttggggttaagt 1046  
 Db 343 ACGGCGGACACAGGTGGTGCATGGCTGCTGTCAGCTCGTGAAGGTGTTGGGTTAAGT 284  
 Qy 1047 ccgcacacgacgcaacccctatcgtattgtgtacc-ttaagttggcactggtacgaa 1105  
 Db 283 CCCGCAACGACGCAACCCCTCGTCTTTAGTTGCCACCGTTGAGTTTGGAAACCCCTGAACAG 224  
 Qy 1106 actgccggtgacaaacccggagggaaggcggtgacgtcctcaatcctcatggcctttatgt 1165  
 Db 223 ACTCCGCTGATAGCCGGAGGAAGGTGAGGATGACGTCAAGTCATCATGTCGCCCTTATGC 164  
 Qy 1166 ccagggccacacgctgtctacaatggccgatacagaggggtgcgaactcgcaagagggag 1225  
 Db 163 CCTGGGCGACACAGTGTCTAAATGGCGGGGACAAAGGGTGCAGTCCCGCGAGGGTGAG 104  
 Qy 1226 ctaattctctaaagtccggtcccgattccggaattgggggtctgcaactcgaccctatgaagtc 1285  
 Db 103 CTAACCTCAAAAACCCCGTCCCTCAGTTCGGATTGAGGCTGCACTAGCCTGCATGAAGCC 44  
 Qy 1286 ggaatcgtcgtagtaatcgcgatacag-catgccgcggtgaatac 1327  
 Db 43 GGAATCGCTAGTAATCCCGGTGACGCATACGGCGGTGAATTC 1



QY	1130	ggcggggatgacgtcaaatctctcatg9ccttctatgtgccaggccacacacagtgctcaaat	1189
Db	301	GGTGAGGATGAGCTCAAGTTCATCATGCCCTTATTCCTCTGGCGACACACAGTCTCAAAAT	360
QY	1190	ggcgcatacagagggtcgcgaactcgcaagggagcttaattctctaaagtgcgtcccg	1249
Db	361	GGCGGGACAAGGGTCCGGATCCCGCAGGGTGAGCTAACCCCAAAAACCCGTCCTCAG	420
QY	1250	ttcggattggggtctgcgaactgacccccatgaatcggaatcgctagtaatcgcgatca	1309
Db	421	TTTCGGATTTCAGAGCTGCACACTCGCTGCATGAAGCCGGAATCGCTACTAATCCCGGTCA	480
QY	1310	g-cataccgcggtgaatcagttcccgaccctgtacacacgcccgctcacacacactgag	1368
Db	481	GCCTACAGCGGGTGAATTCGTTTCCCGGCCCTGTACACACGCCCGCTACACATATGGGAG	540
QY	1369	tggggagcaccggaagtgtcttgttaacgtaaggagacagactaactaagtgaaact	1428
Db	541	CTGGCCATGCCGAAGTCGT-TACCTTAACCGAAGGAGGGGATGCCGAAGCAGGCGCT	599
QY	1429	cgtaagggggtgaagtcgtaacaaagggtaccgtaaatoga	1468
Db	600	AGTGACTGGAGTGAAGTCGTAAACAAGGTAGCCGTACTGGA	639
RESULT 12			
AC019017/c			
LOCUS			
DEFINITION		616 bp DNA	GSS
LOCUS		Homo sapiens genomic DNA, 21q region, clone: B753B2BB067F, genomic survey sequence.	10-OCT-1999
ACCESSION		AG019017	
VERSION		AG019017.1	GI:6017503
KEYWORDS		GSS.	
SOURCE		Homo sapiens DNA, clone: B753B2BB067F.	
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE		1 (bases 1 to 616)	
AUTHORS		Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.	
TITLE		Homo sapiens genomic DNA, chromosome 21q	
JOURNAL		Published only in Database (1999) In press	
REFERENCE		2 (bases 1 to 616)	
AUTHORS		Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.	
TITLE		Direct Submission	
JOURNAL		Submitted (10-OCT-1999) to the DDBJ/EMBL/Genbank databases.	
Masahira Hattori, RIKEN Genomic Sciences Center(GSC) c/o Kitasato University, 1-15-1 Kitasato, Sagamihara 228-8555, Japan [E-mail:hattori@gsc.riks.ac.jp, tel:042-778-9923, Fax:042-778-9924]			
FEATURES			
Location/Qualifiers			
1..616			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="21"			
/clone="B753B2BB067F"			
/map="21q"			
BASE COUNT		131 a 187 c 145 g 153 t	
ORIGIN			
Query Match 23.7%; Score 350.6; DB 190; Length 616;			
Best Local Similarity 78.9%; Pred. No. 2.8e-100;			
Matches 468; Conservative 0; Mismatches 119; Indels 6; Gaps			
QY	872	aattgacgggggtccgcacaaacggtgagcagctggttttaatttcagatgataccacaaac	931
Db	616	AATTGACGGGGGCCCGACACAAACCGTGCACCATGTGTTTAAATTCGATCCACCCGAGA	557
QY	932	acctcacctggggtgacatgcatctgaaatcat-gtagagatatatgagccttcggg--c	988
Db	556	ACCTTACCTGGCTTTCAGATCCACGGAAGTTTTTCAGATGAGATGTGCTTCGGGAAC	497

QY	989	agattcacaagtgctgcattggtgctgctcagctcgtgctgagatgctgggtttaaagctc	1048
Db	496	CGTGCAGACAGCTGCGATGGCTGTCGTGAGCTCGTGTGCTGAAATGTTGGGTTAAGTCC	437
QY	1049	cgcacagcgcgaacccctatcgtatgttctacc--ttaagttgggcactgggtacgaaa	1106
Db	436	CGCAGCAGGCGCAACCTTTATCCTTGTGTCACGGGTCCGGCGGGGAATCAAAGGAGA	377
QY	1107	ctgccggtgcaaacccggaggaaggcgggatgacgtcctaatoctcatggcctttatgtc	1166
Db	376	CTGCCAGTGATAAACTGGAGGAAGGTGGGATGACGTCAAGTCAATCATATGCGCCTTAGCAC	317
QY	1167	cagggccacacagctgctcaaatggcgcatacagagaggtcgccaactcgcaaggaagc	1226
Db	316	CAGGGCTACACAGCTGCTCAATAGGCCCATACAAGAGAACGCGACCTCGCGAGACCAAGC	257
QY	1227	taatctaaagtcggtccagtcgagattgggtctgcaactcgaccccatgaaagtcg	1286
Db	256	GGACCTCATAAAGTGGCTGCTAGTCCGGATTGGAGTCTGCAACTCGACTCATCAAGTCG	197
QY	1287	gaatcgtagttaatcgcggatcagcatcgccggtggaatacgtcccgacacctgtgtacac	1346
Db	196	GAATCGCTAGTAATCGTGSATAGATAGATGCCACGGTGAATACGTTCCGGGGCTTGTATAC	137
QY	1347	accgcccgtcacaccactgagtgggagcaccgcgaagtgctttgtaaccgtaaagga	1406
Db	136	ACCGCCGCTCACACCATGGGAGTGGTTGCCAAAGAAAGTAG-GTAGCTTAACTTCGCGGA	78
QY	1407	gacagactaaggtgaaactcgttaagggggtggaagtcgttaacaaggtacc	1459
Db	77	GGCGCTTACCACCTTTCTGATTTCATGACTGGGTGAAGTCGTAACAAGGTAAAC	25
RESULT 13			
LOCUS	AO956297	675 bp	DNA
DEFINITION	LERAJ197F LERA Arabidopsis thaliana genomic clone LERAJ19, DNA sequence.		
ACCESSION	AO956297		
VERSION	AO956297.1	GI:6783998	
KEYWORDS	GSS.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 675) Buell C.R., Lin X., Pai, G., Barnstead M., Bowman, C., Utterbach, T., Feldblum, T., Liang, F., Creasy, T. and Fraser, C.M.		
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atetigr.org For additional information, see <a href="http://www.tigr.org/tdb/at/at.html">http://www.tigr.org/tdb/at/at.html</a> Similar to A. thaliana chloroplast sequence (GB:AF000423) Seq primer: TF Class: shotgun.		
FEATURES	Location/Qualifiers		
source	1..675 /organism="Arabidopsis thaliana" /strain="Landsberg erecta" /db_xref="taxon:3702" /clone_lib="LERA" /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."		



```
BASE COUNT      176 a      153 c      222 g      124 t
ORIGIN

Query Match      23.5%; Score 346.8; DB 165; Length 675;
Best Local Similarity 73.8%; Pred. No. 4.7e-99;
Matches 484; Conservative 0; Mismatches 162; Indels 10; Gaps 3;

Qy 190 gcttggagatgagcccgccgagcttagctagttggtgagtaalgtctccaccaggcga 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28 GCCCGAGGAGGGCGTCCGCTGATAGTGTAGTGTGGTGGAGCAATAGCTTACCAAGGCGA 87

Qy 250 cgatcgtagccgctgagaggtgtccgccaataatgaactgagacagcgtccatcac 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88 TGATCAGTAGTGTGTCCGAGAGGATGATCAGCCACTGGGACTGAGACACGGCCGACAG 147

Qy 310 tcttacggagcgagcagcttaagaatcttctcaatggggaacccctgaagcagcgacg 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 TCCTACGGGAGCGACGAGTGGGGAATTTCCGCAATGGCGCAAGCCGTGACGGAGCAATG 207

Qy 370 ccgctgaaacaaagaggtcttcggtattgtaagtctataggcaggaaataaagcagc 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 CCGCGTGGAGGTAGAAGGCCCTACGGGTCTCTGAACCTCTTTTCCCGAG--AGAAGAAGCAAT 265

Qy 430 aatgtgatgtgacctgcctaaagcagcgcttaactacgtccagcagccgagcgtgaat 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 GACGGTATCTGGGA-----ATAGCATCGGCTTAACCTCTGTGCCAGCAGCGCGGTAAAT 319

Qy 490 acgtatggtgcaagctgttctcgaatacattggcgtaaaagggtgcgtagcgatttgt 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 ACAGAGGATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCTGTAGTGGCTTTT 379

Qy 550 aagtcagggtgtaaaactgcgggtcaaccgtgacctggaacttgaactcaaatctctg 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 AAGTCCCGCGCTCAATCCAGGGCTCAACCTCGACAGCGGTGGAACTACCAAGCTTG 439

Qy 610 agtttgagaggaagaagtcaggtgtagcgtggaatcgtagatatcttgagg 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 440 AGTACGGTAGGGGACAGGGGAATTTCCGGTGGCGGTTGAATTCGTAGAGATCGGAAG 499

Qy 670 aacaccagtgcgaaggcagctgtggtctcaaaactgacgctgaggcacgaagcgtgg 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500 AACACCAACGGCGAAGCACACTCTCTGGCGCGACACTGCACACTGAGAGACCAAGCTAGG 559

Qy 730 gtatgaacagggtatagatacccggttaatccagccctaaactgttctaccagttgtt 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 560 GGAGCGAATGGGATTAGATACCCCGAGTAGTCTTAGCCGTAACAGTGGATAGTGGCGCT 619

Qy 790 gggggttttaaccc--tcagttaacgaaccttaacgattaaagtacagccctgggga 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 620 GTGGGTATCGACCGTGCAGTGTGTAGCTAACGGTTAAGTATCCCGCCCTGGGGA 675

RESULT 14
AW833978      659 bp      mRNA      EST      18-MAY-2000
LOCUS
DEFINITION QV0-TT0009-191199-056-b02 TT0009 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW833978
VERSION AW833978.1 GI:7927952
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 659)
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Goldman,G.H., Carvalho,A.P., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

FEATURES
source
1..659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TT0009"
/dev_stage="Adult"
/note="Organ: testis; Vector: puc18; Site:1; Site:2;
Sma1: A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      152 a      146 c      234 g      126 t      1 others
ORIGIN

Query Match      23.3%; Score 344.; DB 95; Length 659;
Best Local Similarity 74.4%; Pred. No. 3.6e-98;
Matches 490; Conservative 0; Mismatches 156; Indels 13; Gaps 4;

Qy 254 cggtagcggcctgagaggtgtccggccacaatggaactgagacacggtccatactct 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13 CGGTGTCGTGGTCTGAGAGGATGATCAG-CACACTGGGACTGAGACACGCCCGCAGACTCT 71

Qy 314 acggagagcagcagtttaagaatcttgcctcaatggggaaacccctgaagcagcagccgc 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 ACGGAGGACGACGATGGGGAATATTGGACAATGGGGGCAACCCCTGATCCAGCAATGCCGC 131

Qy 374 gtgaacgaagaaggtcttcggattgttaagtccaatagcaggaaaaataaagcagaatg 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 CTGACTGAAGAAGGCCCTTAGGGTTGTAAGCTCTTTCCGCCGTGACGATGATGACGGTAA 191

Qy 434 tgatgatggtacctgcctaaagcaccggctaactacgtgcccagcagccggtataacgt 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 CGGGGAGAG-----AAGCCCGGCTAATTCGTGCCAGCAGCGCGGTAATACGG 241

Qy 494 atggtgcaagcgttctcggaatcattggcgctaaaggtgcgttagcgagatttgaagt 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 AAGGGCTAGCGTGTTCGGAATTTACTGGCGTAAGCGCATGTAGCGCGCGGATCAAGT 301

Qy 554 caggtgtgaaaaactgcgggctcaaccctgtgacctgtgaaactacaagtcttgagtt 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 GGGCGTCAAAAGGCCCTGGCTCAACCTGGGAACCTGCGCTTCAGACTGTTTCTCGAG-T 360

Qy 614 tggagagagcaagtggaaattccaggtgtagcggtagaaatcgtagatcttggagaaca 673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 TCGGAGAGGGTGGTGAATTTCCAGTGTAGAGGTGAATTCGTAGATATTGGGAGGAACA 420

Qy 674 ccagtggcgaaggagcacttgctgctcaaaactgaactgacgtgagcagcaagcgtgggtag 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 CCGGTGGCAGAGGGCGCCACCTGGCCCGAGACTGACGCTCAAGTCCGAAAGCGTGGGAG 480

Qy 734 taaacggattagatcaccccggttaatccacccctaaacgttgtctaccagttgtggg 793
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 CAACAGGATTAGATACCTCGTAGTCCACGCTGTAAACGATGGGTGTAGATGTCGGGG 540
```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:22:56 ; Search time 207.51 Seconds  
(without alignments)  
39,827 Million cell updates/sec

Title: US-09-380-826A-2  
Perfect score: 22  
Sequence: 1 ttttgatcacagatttgata 22

Scoring table:  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

1:	/cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	19 V58901	Leptospiira rRNA ge
2	22	100.0	22	19 V58897	L. fainei nucleoti
3	22	100.0	1477	19 V58896	L. fainei nucleoti
c 4	16.8	76.4	2582	18 T90543	Maize male fertili
c 5	16.2	73.6	951	20 V84598	Human secreted pro
c 6	16.2	73.6	1291	19 V59636	Human secreted pro
c 7	16.2	73.6	1473	20 X61757	B. burgdorferi anc
8	16.2	73.6	2000	17 T17111	Acetyl-CoA-carboxy
9	16.2	73.6	2001	17 T39904	Maize acetyl CoA c
10	16.2	73.6	2001	21 Z49816	ECORI fragment of
11	16.2	73.6	3822	18 T64683	M. leprae gyrA pre
12	16.2	73.6	4345	14 Q42933	A3 maize ACCase cd

13	16.2	73.6	7470	17 T39905
14	16.2	73.6	7470	19 V29317
15	16.2	73.6	7470	21 Z49820
c 16	16.2	73.6	10811	19 V59091
17	15.8	71.8	843	20 Z15467
18	15.8	71.8	961	19 X14095
19	15.8	71.8	4146	21 Z51556
c 20	15.6	70.9	297	20 V88753
c 21	15.6	70.9	795	20 V72025
c 22	15.6	70.9	834	20 V72026
c 23	15.6	70.9	1075	18 T72715
c 24	15.6	70.9	1096	13 Q20498
c 25	15.6	70.9	1194	20 V72027
c 26	15.6	70.9	1560	18 T49864
c 27	15.6	70.9	1584	19 V28845
c 28	15.6	70.9	1761	21 Z45339
c 29	15.6	70.9	1989	20 V81481
c 30	15.6	70.9	2354	19 V50429
c 31	15.6	70.9	2651	20 X13369
c 32	15.6	70.9	7745	18 T79859
c 33	15.6	70.9	7745	18 T59308
c 34	15.6	70.9	7745	20 X04183
c 35	15.6	70.9	9824	14 Q40763
c 36	15.2	69.1	349	13 Q34809
c 37	15.2	69.1	493	20 X13759
c 38	15.2	69.1	1028	17 T36923
c 39	15.2	69.1	1848	16 Q86996
c 40	15.2	69.1	1848	16 Q86998
c 41	15.2	69.1	2202	17 T36922
c 42	15.2	69.1	2361	20 X04316
c 43	15.2	69.1	2598	18 T72286
c 44	15.2	69.1	2598	21 Z98878
c 45	15.2	69.1	5420	12 Q11643

## ALIGNMENTS

RESULT 1  
V58901  
ID V58901 standard; DNA; 22 BP.  
XX  
AC V58901;  
XX  
DT 20-JAN-1999 (first entry)  
XX  
DE Leptospiira rRNA gene nucleotide sequence.  
XX  
KW Infection; pathogenic Leptospiira; protective immunity; therapy;  
XX diagnosis; SS.  
OS Leptospiira sp.  
XX  
FN WO9840099-A1.  
XX  
PD 17-SEP-1998.  
XX  
PF 06-MAR-1998; 98WO-AU00145.  
XX  
PR 07-MAR-1997; 97AU-0005494.  
XX  
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
XX (PIGR-) PIG RES & DEV CORP.  
XX  
PI Chappel RJ;  
XX  
DR WPI; 1998-520791/44.  
XX  
PT New isolated pathogenic Leptospiira bacterium - useful for, e.g  
PT developing products for conferring protective immunity, and for  
PT prophylactic or therapeutic treatment  
XX  
PS Claim 15; Page 72; 94pp; English.

Maize acetyl CoA c  
Maize ACCase enzym  
Maize acetyl CoA c  
Potato pollen cell  
Human gene express  
H. pylori GHPO 131  
Human hypoxia resp  
EST clone HK650.  
Adenovirus PACTSG2  
Adenovirus SCAR.RG  
C. elegans inhibit  
Encodes fibrinogen  
Adenovirus PACSG2S  
Callosellasma rhod  
Human coxsackievir  
DNA encoding a Bac  
Black Creek Canal  
Human coxsackievir  
Enterococcus faeca  
Genomic DNA encodi  
SH2-mRev6 gene (s  
Shrunken-2 gene va  
Wild-type shrunken  
Human type IV coll  
Enterococcus faeca  
Human OVCA2 tumour  
Aminopeptidase O12  
Aminopeptidase O12  
Human OVCA1 tumour  
Human secreted pro  
Breast cancer mamm  
MMTV env protein n  
Partial human comp

XX This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 CC L. fainei. The LS bacteria can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.047;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttgatcacacaagatttgata 22  
 Db 1 tgttgatcacacaagatttgata 22

RESULT 2  
 V58897  
 ID V58897 standard; DNA; 22 BP.  
 AC V58897;  
 DT 20-JAN-1999 (first entry)  
 DE L. fainei nucleotide sequence.  
 XX  
 KW Infection; pathogenic Leptospira; protective immunity; therapy;  
 diagnosis; ss.  
 OS Leptospira fainei.  
 PN WO9840099-A1.  
 XX  
 PD 17-SEP-1998.  
 PF 06-MAR-1998; 98WO-AU00145.  
 PR 07-MAR-1997; 97AU-0005494.  
 XX  
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 (PIGR-) PIG RES & DEV CORP.  
 PI Chappel RJ;  
 XX  
 DR WPI; 1998-520791/44.  
 XX  
 PT New isolated pathogenic Leptospira bacterium - useful for, e.g  
 developing products for conferring protective immunity, and for  
 PT prophylactic or therapeutic treatment  
 XX  
 PS Claim 15; Page 70; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 CC L. fainei. The LS bacteria can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.047;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttgatcacacaagatttgata 22  
 Db 1 tgttgatcacacaagatttgata 22

RESULT 3  
 V58896  
 ID V58896 standard; DNA; 1477 BP.  
 AC V58896;  
 DT 20-JAN-1999 (first entry)  
 DE L. fainei nucleotide sequence.  
 XX  
 KW Infection; pathogenic Leptospira; protective immunity; therapy;  
 diagnosis; ss.  
 OS Leptospira fainei.  
 PN WO9840099-A1.  
 XX  
 PD 17-SEP-1998.  
 PF 06-MAR-1998; 98WO-AU00145.  
 PR 07-MAR-1997; 97AU-0005494.  
 XX  
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 (PIGR-) PIG RES & DEV CORP.  
 PI Chappel RJ;  
 XX  
 DR WPI; 1998-520791/44.  
 XX  
 PT New isolated pathogenic Leptospira bacterium - useful for, e.g  
 developing products for conferring protective immunity, and for  
 PT prophylactic or therapeutic treatment  
 XX  
 PS Claim 15; Page 69-70; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 CC L. fainei. The LS bacteria can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 SQ Sequence 1477 BP; 390 A; 334 C; 439 G; 314 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 1477;  
 Best Local Similarity 100.0%; Pred. No. 0.072;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttgatcacacaagatttgata 22  
 Db 154 tgttgatcacacaagatttgata 175

RESULT 4  
 T90543/c  
 ID T90543 standard; DNA; 2582 BP.  
 AC T90543;  
 DT 13-FEB-1998 (first entry)  
 DE Maize male fertility gene 233 Zm41-A.  
 XX

KW Zm41-A; male fertility; hybrid seed; transgenic plant; antisense;  
 KW ribozyme; male sterile; maize; Ms41-A; ds.

OS Zea mays.

XX Key Location/Qualifiers

XX exon 823..915

FT intron /\*tag= a

FT exon 916..935

FT exon /\*tag= b

FT exon 936..1090

FT intron /\*tag= c

FT intron 1091..1167

FT exon /\*tag= d

FT exon 1168..1332

FT exon /\*tag= e

XX WO9723618-A1.

PN 03-JUL-1997.

PD 20-DEC-1996; 96WO-GB03191.

XX 21-DEC-1995; 95GB-0026218.

XX (GENE-) GENE SHEARS PTY LTD.

XX Baudot G, Garcia D, Hodge R, Perez P;

XX WPI; 1997-351055/32.

XX Nucleic acid encoding proteins involved in male fertility in plants

PT - used to control fertility and for production of hybrid seed

XX Claim 3; Fig 15; 85pp; English.

XX This DNA sequence comprises the maize Z33 Zm41-A gene. This is an

CC orthologue of Arabidopsis Ms41-A (see T90522), a gene that confers

CC male fertility. It was obtained following a database search for

CC sequences that showed homology to Ms41-A DNA, isolate of a

CC partial cDNA clone, and use of this clone to isolate Zm41-A genes

CC Z31 (T90542), Z33 (T90543) and Z35 (T90544) from genomic lambda

CC libraries. The genes show a high level of conservation. Z35 may

CC be derived from Z31 via genetic rearrangements, deletions and/or

CC insertions. Z33 has subsequent deletions from Z35 and is

CC truncated, having only exons 3, 5 and 6. The Ms41-A and Zm41-A

CC genes, antisense or ribozyme sequences can be used to produce

CC transgenic plants with controlled male fertility. Male sterile

CC plants are useful for hybrid seed production, particularly in

CC plants where restoration of fertility is not needed, e.g.

CC Brassicaceae, lettuce, spinach and onions.

XX SQ Sequence 2582 BP; 670 A; 512 C; 537 G; 863 T; 0 other;

Query Match 76.48; Score 16.8; DB 18; Length 2582;

Best Local Similarity 90.08; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ttgtggatcaccaagatttga 20

||||| ||||| ||||| |||||

Db 268 TGTTCATCACAAGATCTGA 249

RESULT 5

ID V84598/c

XX V84598 standard; DNA; 951 BP.

AC V84598;

XX 01-MAR-1999 (first entry)

DT Human secreted protein gene 188 clone HHPSP70.

XX

DE

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 XX Homo sapiens.  
 XX WO9854963-A2.  
 PN 10-DEC-1998.  
 PD 04-JUN-1998; 98WO-US11422.  
 PF 18-DEC-1997; 97US-0070923.  
 XX 06-JUN-1997; 97US-0048877.  
 XX 06-JUN-1997; 97US-0048881.  
 XX 06-JUN-1997; 97US-0048884.  
 XX 06-JUN-1997; 97US-0048893.  
 XX 06-JUN-1997; 97US-0048896.  
 XX 06-JUN-1997; 97US-0048899.  
 XX 06-JUN-1997; 97US-0048915.  
 XX 06-JUN-1997; 97US-0048949.  
 XX 06-JUN-1997; 97US-0048964.  
 XX 06-JUN-1997; 97US-0048972.  
 XX 06-JUN-1997; 97US-0049020.  
 XX 06-JUN-1997; 97US-0049375.  
 XX 05-SEP-1997; 97US-0057628.  
 XX 05-SEP-1997; 97US-0057635.  
 XX 05-SEP-1997; 97US-0057644.  
 XX 05-SEP-1997; 97US-0057647.  
 XX 05-SEP-1997; 97US-0057650.  
 XX 05-SEP-1997; 97US-0057661.  
 XX 05-SEP-1997; 97US-0057667.  
 XX 05-SEP-1997; 97US-0057761.  
 XX 05-SEP-1997; 97US-0057764.  
 XX 05-SEP-1997; 97US-0057770.  
 XX 05-SEP-1997; 97US-0057775.  
 XX 05-SEP-1997; 97US-0057778.  
 XX 06-JUN-1997; 97US-0048875.  
 XX 06-JUN-1997; 97US-0048878.  
 XX 06-JUN-1997; 97US-0048882.  
 XX 06-JUN-1997; 97US-0048885.  
 XX 06-JUN-1997; 97US-0048894.  
 XX 06-JUN-1997; 97US-0048897.  
 XX 06-JUN-1997; 97US-0048900.  
 XX 06-JUN-1997; 97US-0048916.  
 XX 06-JUN-1997; 97US-0048962.  
 XX 06-JUN-1997; 97US-0048970.  
 XX 06-JUN-1997; 97US-0048974.  
 XX 06-JUN-1997; 97US-0049373.  
 XX 05-SEP-1997; 97US-0057584.  
 XX 05-SEP-1997; 97US-0057629.  
 XX 05-SEP-1997; 97US-0057642.  
 XX 05-SEP-1997; 97US-0057645.  
 XX 05-SEP-1997; 97US-0057648.  
 XX 05-SEP-1997; 97US-0057651.  
 XX 05-SEP-1997; 97US-0057668.  
 XX 05-SEP-1997; 97US-0057668.  
 XX 05-SEP-1997; 97US-0057762.  
 XX 05-SEP-1997; 97US-0057765.  
 XX 05-SEP-1997; 97US-0057771.  
 XX 05-SEP-1997; 97US-0057776.  
 XX 06-JUN-1997; 97US-0048876.  
 XX 06-JUN-1997; 97US-0048880.  
 XX 06-JUN-1997; 97US-0048883.  
 XX 06-JUN-1997; 97US-0048892.  
 XX 06-JUN-1997; 97US-0048895.  
 XX 06-JUN-1997; 97US-0048898.

us-09-380-826a-2.rng

Wed Mar 28 14:03:15 2001

PR 06-JUN-1997; 97US-0048901.  
 PR 06-JUN-1997; 97US-0048917.  
 PR 06-JUN-1997; 97US-0048963.  
 PR 06-JUN-1997; 97US-0048971.  
 PR 06-JUN-1997; 97US-0049019.  
 PR 06-JUN-1997; 97US-0049374.  
 PR 06-JUN-1997; 97US-0057627.  
 PR 05-SEP-1997; 97US-0057634.  
 PR 05-SEP-1997; 97US-0057643.  
 PR 05-SEP-1997; 97US-0057646.  
 PR 05-SEP-1997; 97US-0057649.  
 PR 05-SEP-1997; 97US-0057654.  
 PR 05-SEP-1997; 97US-0057666.  
 PR 05-SEP-1997; 97US-0057760.  
 PR 05-SEP-1997; 97US-0057763.  
 PR 05-SEP-1997; 97US-0057769.  
 PR 05-SEP-1997; 97US-0057774.  
 PR 05-SEP-1997; 97US-0057777.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;  
 XX Fan P, Feng P, Ferris AM, Fischer CL, Florence C;  
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;  
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;  
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;  
 XX  
 XX WPI: 1999-059865/05.  
 DR P-PSDB; W88721.  
 DR  
 XX New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PT  
 XX  
 XX Claim 4: Page 450-451; 772pp: English.  
 PS  
 XX The invention relates to nucleic acid sequences (V84411 to V84633)  
 CC encoding human secreted proteins (W88534 to W88756). The secreted protein  
 CC gene sequences are deposited with the ATCC under deposit numbers ATCC  
 CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
 CC cells comprising recombinant vectors containing the nucleic acid  
 CC sequences are used for the recombinant production of the secreted  
 CC proteins. The polynucleotide and amino acid sequences are useful for  
 CC by determining the amount of the new polypeptides in a sample or by  
 CC useful for preventing, treating or ameliorating medical conditions e.g.  
 CC by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the polynucleotides, based on  
 CC which tissues they are most highly expressed in, and include developing  
 CC products for the diagnosis or treatment of cancer, neurodegenerative  
 CC disorders, developmental abnormalities and foetal deficiencies, blood  
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
 CC ischemic shock, Alzheimer's and cognitive disorders involving osteoclasts  
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts  
 CC such as osteoporosis, arthritis or malignancies, infections of testes, lung  
 CC or thymus digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The present sequence represents a gene encoding a human secreted protein  
 CC (see descriptor line for gene number and clone identification).  
 CC  
 XX Sequence 951 BP; 296 A; 134 C; 192 G; 329 T; 0 other;  
 SQ

Query Match 73.6%; Score 16.2; DB 20; Length 951;  
 Best Local Similarity 85.7%; Pred. No. 47;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgttgatcacagattgat 21  
 ||||| ||||| |||||  
 DB 524 TGTGGTCTCAAGATCTCAT 504

RESULT 6  
 V59636/c  
 ID V59636 standard; DNA; 1291 BP.  
 XX  
 AC V59636;  
 XX  
 DT 19-JAN-1999 (first entry)  
 XX  
 XX Human secreted protein gene 126 clone HELGH31.  
 DE  
 XX  
 XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09839448-A2.  
 XX  
 PD 11-SEP-1998.  
 XX  
 PF 06-MAR-1998; 98WO-US04493.  
 XX  
 XX 02-OCT-1997; 97US-0061060.  
 PR 07-MAR-1997; 97US-0038621.  
 PR 07-MAR-1997; 97US-0040161.  
 PR 07-MAR-1997; 97US-0040162.  
 PR 07-MAR-1997; 97US-0040163.  
 PR 07-MAR-1997; 97US-0040333.  
 PR 07-MAR-1997; 97US-0040334.  
 PR 07-MAR-1997; 97US-0040336.  
 PR 11-APR-1997; 97US-0040626.  
 PR 11-APR-1997; 97US-0043311.  
 PR 11-APR-1997; 97US-0043312.  
 PR 11-APR-1997; 97US-0043313.  
 PR 11-APR-1997; 97US-0043314.  
 PR 11-APR-1997; 97US-0043568.  
 PR 11-APR-1997; 97US-0043569.  
 PR 11-APR-1997; 97US-0043576.  
 PR 11-APR-1997; 97US-0043578.  
 PR 11-APR-1997; 97US-0043580.  
 PR 11-APR-1997; 97US-0043669.  
 PR 11-APR-1997; 97US-0043670.  
 PR 11-APR-1997; 97US-0043671.  
 PR 11-APR-1997; 97US-0043672.  
 PR 11-APR-1997; 97US-0043674.  
 PR 23-MAY-1997; 97US-0047492.  
 PR 23-MAY-1997; 97US-0047500.  
 PR 23-MAY-1997; 97US-0047501.  
 PR 23-MAY-1997; 97US-0047502.  
 PR 23-MAY-1997; 97US-0047503.  
 PR 23-MAY-1997; 97US-0047581.  
 PR 23-MAY-1997; 97US-0047582.  
 PR 23-MAY-1997; 97US-0047583.  
 PR 23-MAY-1997; 97US-0047584.  
 PR 23-MAY-1997; 97US-0047585.  
 PR 23-MAY-1997; 97US-0047586.  
 PR 23-MAY-1997; 97US-0047587.  
 PR 23-MAY-1997; 97US-0047588.  
 PR 23-MAY-1997; 97US-0047589.  
 PR 23-MAY-1997; 97US-0047590.  
 PR 23-MAY-1997; 97US-0047592.  
 PR 23-MAY-1997; 97US-0047593.  
 PR 23-MAY-1997; 97US-0047594.  
 PR 23-MAY-1997; 97US-0047595.  
 PR 23-MAY-1997; 97US-0047596.  
 PR 23-MAY-1997; 97US-0047597.  
 PR 23-MAY-1997; 97US-0047598.



PR 23-MAY-1997; 97US-0047599.  
 PR 23-MAY-1997; 97US-0047600.  
 PR 23-MAY-1997; 97US-0047601.  
 PR 23-MAY-1997; 97US-0047612.  
 PR 23-MAY-1997; 97US-0047613.  
 PR 23-MAY-1997; 97US-0047614.  
 PR 23-MAY-1997; 97US-0047615.  
 PR 23-MAY-1997; 97US-0047617.  
 PR 23-MAY-1997; 97US-0047618.  
 PR 23-MAY-1997; 97US-0047632.  
 PR 23-MAY-1997; 97US-0047633.  
 PR 06-JUN-1997; 97US-0048964.  
 PR 06-JUN-1997; 97US-0048974.  
 PR 13-JUN-1997; 97US-0049610.  
 PR 08-JUL-1997; 97US-0051926.  
 PR 16-JUL-1997; 97US-0052874.  
 PR 18-AUG-1997; 97US-0055724.  
 PR 22-AUG-1997; 97US-0056630.  
 PR 22-AUG-1997; 97US-0056631.  
 PR 22-AUG-1997; 97US-0056632.  
 PR 22-AUG-1997; 97US-0056636.  
 PR 22-AUG-1997; 97US-0056637.  
 PR 22-AUG-1997; 97US-0056662.  
 PR 22-AUG-1997; 97US-0056664.  
 PR 22-AUG-1997; 97US-0056845.  
 PR 22-AUG-1997; 97US-0056862.  
 PR 22-AUG-1997; 97US-0056864.  
 PR 22-AUG-1997; 97US-0056872.  
 PR 22-AUG-1997; 97US-0056874.  
 PR 22-AUG-1997; 97US-0056875.  
 PR 22-AUG-1997; 97US-0056876.  
 PR 22-AUG-1997; 97US-0056877.  
 PR 22-AUG-1997; 97US-0056878.  
 PR 22-AUG-1997; 97US-0056879.  
 PR 22-AUG-1997; 97US-0056880.  
 PR 22-AUG-1997; 97US-0056881.  
 PR 22-AUG-1997; 97US-0056882.  
 PR 22-AUG-1997; 97US-0056884.  
 PR 22-AUG-1997; 97US-0056886.  
 PR 22-AUG-1997; 97US-0056887.  
 PR 22-AUG-1997; 97US-0056888.  
 PR 22-AUG-1997; 97US-0056889.  
 PR 22-AUG-1997; 97US-0056892.  
 PR 22-AUG-1997; 97US-0056893.  
 PR 22-AUG-1997; 97US-0056894.  
 PR 22-AUG-1997; 97US-0056903.  
 PR 22-AUG-1997; 97US-0056908.  
 PR 22-AUG-1997; 97US-0056909.  
 PR 22-AUG-1997; 97US-0056910.  
 PR 22-AUG-1997; 97US-0056911.  
 PR 05-SEP-1997; 97US-0057650.  
 PR 05-SEP-1997; 97US-0057659.  
 PR 05-SEP-1997; 97US-0057761.  
 PR 12-SEP-1997; 97US-0058785.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PA Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress CA;  
 PI Feng P, Fertie AM, Fischer CL, Florence KA, Greene JM, Hu JS;  
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 XX WPI; 1998-506364/43.  
 DR P-PSDB; W74854.  
 XX  
 XX New isolated human genes and the secreted polypeptide(s) they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation of blood disorders  
 XX  
 PS Claim 1; Page 360-361; 721pp; English.  
 CC  
 CC This sequence represents a nucleic acid molecule designated Gene 126 from  
 CC the human CDNA clone HELGH31 (deposited as clone ATCC 97902 and ATCC

CC 209048) which encodes a secreted human protein. The gene can be used to  
 CC generate fusion proteins by linking to the gene to a human immunoglobulin  
 CC Fc portion (e.g. V59502) for increasing the stability of the fused  
 CC protein as compared to the human protein only.  
 CC The invention relates to 186 novel genes and their fragments (nucleic  
 CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also, pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the 186 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see V59511 for described  
 CC uses).  
 XX  
 SQ Sequence 1291 BP; 391 A; 228 C; 203 G; 469 T; 0 other;  
 Query Match 73.6%; Score 16.2; DB 19; Length 1291;  
 Best Local Similarity 85.7%; Pred. No. 48;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 tgttggtacacagatttgat 21  
 ||| ||||| |||||  
 Db 367 TGTATGATCACATGATTGAT 347  
 RESULT 7  
 X61757/C  
 ID X61757 standard; DNA; 1473 BP.  
 XX  
 XX X61757;  
 AC  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein coding sequence, f893.nt.  
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.  
 OS Borrelia burgdorferi.  
 XX  
 PN WO9859071-A1.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-US12718.  
 XX  
 PR 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMMUNE INC.  
 XX  
 PI Choi GH, Erwin AL, Hanson MS, Lathigra R;  
 XX  
 DR WPI; 1999-189980/16.  
 DR P-PSDB; Y20060.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by Borrelia, particularly Lyme disease  
 XX  
 XX Claim 1; Page 180; 275pp; English.  
 PS  
 XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.

us-09-380-826a-2.rng

Wed Mar 28 14:03:15 2001

CC promoter, e.g. a cauliflower-mosaic virus 35S promoter or  
 CC nopaline-synthase promoter, and N-terminal maize chloroplast  
 CC transit peptide sequence in a plasmid vector for high-level  
 CC tissue-specific gene expression, to confer herbicide resistance  
 CC and/or alter the oil content of the plant, or to increase the plant  
 CC ACCase yield, so that the crystallized enzyme may be used to screen  
 CC and identify other herbicides that bind to and inhibit the enzyme.

SQ Sequence 1473 BP; 622 A; 232 C; 121 G; 498 T; 0 other;  
 Query Match 73.6%; Score 16.2; DB 20; Length 1473;  
 Best Local Similarity 85.7%; Pred. No. 49;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gttggtacacaaagtattgata 22  
 ||||| ||||| ||||| |||||  
 Db 68 GCTGCATCACAGGATTGATA 48

XX Sequence 2000 BP; 546 A; 406 C; 494 G; 554 T; 0 other;

Query Match 73.6%; Score 16.2; DB 17; Length 2000;  
 Best Local Similarity 85.7%; Pred. No. 51;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gttggtacacaaagtattgata 22  
 ||||| ||||| ||||| |||||  
 Db 42 gttggtacacaaagtattgata 62

RESULT 9  
 T39904  
 ID T39904 standard; CDNA; 2001 BP.

XX T39904;  
 AC T39904;  
 DT 23-JAN-1997 (first entry)  
 DE Maize acetyl CoA carboxylase cDNA clone EcoRI fragment.  
 KW Acetyl CoA carboxylase; ACCase; herbicide tolerance;  
 KW cyclohexanedione; aryloxyphenoxypropionic acid; vegetable oil;  
 KW oilseed; maize; corn; ss.  
 OS Zea mays line A188.  
 PN W09631609-A2.  
 PD 10-OCT-1996.

PF 04-APR-1996; 96WO-US04625.  
 PR 05-APR-1995; 95US-0417089.  
 XX (MINU ) UNIV MINNESOTA.  
 XX Egli MA, Gengenbach BG, Gronwald JW, Lutz SM, Somers DA;  
 PI Wyse DL;  
 XX WPI; 1996-465030/46.  
 DR DNA encoding maize acetyl coenzyme A carboxylase gene - used for  
 PT prodn. of plants with herbicide tolerance or altered oil content  
 PT Claim 2; Page 77; 131pp; English.

XX A 2 kb EcoRI fragment (T39904) of lambda clone #15-14 includes a  
 CC portion of a maize acetyl CoA carboxylase (ACCase) gene located at  
 CC bases 2883 to 83' stop codon. Clone #15-14 was isolated  
 CC from a lambda gt11 cDNA library of maize inbred A188 seedlings by  
 CC screening with maize ACCase antiserum. Other ACCase DNA fragments  
 CC (see also T39906-16) and a complete cDNA (T39905) coding for ACCase  
 CC (W05590) have also been isolated. ACCase DNA can be incorporated  
 CC into a vector and used to increase the herbicide tolerance or oil  
 CC content of a transgenic plant, or used to produce ACCase in  
 CC heterologous systems.

XX Sequence 2001 BP; 546 A; 406 C; 494 G; 554 T; 1 other;

Query Match 73.6%; Score 16.2; DB 17; Length 2001;  
 Best Local Similarity 85.7%; Pred. No. 51;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX 07-JUL-1996 (first entry)  
 DE Acetyl-CoA-carboxylase gene fragment.

XX Maize; acetyl-CoA-carboxylase; fluzifop; sethoxydim; herbicide;  
 KW probe; transcarboxylase active site; biotin binding site;  
 KW screening; cassette; antisense; promoter; transit peptide; plasmid;  
 KW tissue-specific gene expression; herbicide resistance; oilseed;  
 KW oil content; transgenic plant; crop improvement; ss.

XX Zea mays.

XX Key Location/Qualifiers  
 FH misc\_feature 1770..1946  
 FT /\*tag= a  
 FT /product= Transcarboxylase active site region  
 FT misc\_feature 59..70  
 FT /\*tag= b  
 FT /product= Biotin binding site

PN US5498544-A..

PD 12-MAR-1996.

PF 10-NOV-1988; 88US-0269584.

PR 05-FEB-1993; 93US-0014326.

PR 10-NOV-1988; 88US-0269584.

PR 18-JUN-1990; 90US-0538674.

PR 21-JUL-1992; 92US-0917462.

XX (MINU ) UNIV MINNESOTA.

XX Egli MA, Gengenbach BG, Gronwald JW, Lutz SM, Somers DA;  
 XX Wyse DL;  
 XX WPI; 1996-159692/16.

XX Expression cassette containing plant acetyl CoA gene - used to  
 XX confer herbicide tolerance or to increase oil content of plants  
 XX Claim 1; Column 35-38; 31pp; English.

XX The sequence encodes a maize acetyl-CoA-carboxylase (ACCase,  
 CC EC-6.4.1.2) fragment. The enzyme catalyzes the formation of  
 CC malonyl-CoA from acetyl-CoA and bicarbonate, and requires a biotin  
 CC prosthetic group for activity. The enzyme is inhibited by several  
 CC herbicides (e.g. fluzifop and sethoxydim) in monocotyledons. The  
 CC sequence is a 2-kb EcoRI fragment of phage lambda-gt11 clone 15-14,  
 CC isolated from a maize-inbred A188 seedling cDNA library by  
 CC screening with maize ACCase antiserum. Probes generated from the  
 CC transcarboxylase active site domain (T17112) and the biotin binding  
 CC site domain (T17113) coding sequences may be used to screen plants  
 CC for similar genes. An expression cassette containing the gene, a  
 CC mutant or an antisense DNA sequence may be combined with a plant

Oy 2 gttggatcacagaattgata 22  
 ||||| ||||| |||||  
 Db 42 gttggatcacagaattgta 62

## RESULT 10

Z49816  
 ID Z49816 standard; DNA; 2001 BP.

AC XX Z49816;

DT XX 18-APR-2000 (first entry)

DE XX EcoRI fragment of lamda clone #15-14 with maize ACCase gene portion.

KW XX Herbicide resistance; gene modification; lamda clone #15-14;

KW maize acetyl CoA carboxylase; ACCase; herbicide; plant oil content;

KW marker-assisted plant selection; groat oil trait;

KW restriction fragment length polymorphism mapping;

KW high-energy animal feed; low-fat human food; ds.

XX OS Zea mays.

XX PN WO9967367-A1.

XX PD 29-DEC-1999.

XX PF 22-JUN-1999; 99WO-US14022.

XX PR 22-JUN-1998; 98US-0090240.

XX PR 02-JUL-1998; 98US-0091640.

XX PA (MINU ) UNIV MINNESOTA.

PA (USDA ) US DEPT OF AGRICULTURE.

PA (EGLI/) EGLI M A.

PA (GROH/) GROH S.

PA (KIAN/) KIANIAN S F.

PA (PHIL/) PHILLIPS R L.

PA (RINE/) RINES H W.

PA (SOME/) SOMERS D A.

XX PI Egli MA, Groh S, Kianian SF, Phillips RL, Rines HW, Somers DA;

XX DR WPI; 2000-147205/13.

XX New DNA encoding acetyl-CoA carboxylase from oats, used to produce

PT transformed plants with herbicide resistance and altered oil content

PS Disclosure; Fig 10; 197pp; English.

XX The present DNA sequence is a 2kb EcoRI fragment of lamda clone #15-14 including a portion of a maize acetyl CoA carboxylase (ACCase) gene located at bases 2883 to 83 from the 3' stop codon. ACCase is involved in fatty acid synthesis and is the target of the specified herbicides. Transformation of plants with ACCase imparts resistance to cyclohexanedione and aryloxyphenoxypyrrolic acid herbicides and alter the oil content. The ACCase DNA is also used as source of probes and primers for the identification of transgenic plants; in marker-assisted plant selection and for restriction fragment length polymorphism mapping, used for high-energy animal feed and high-fiber, low-fat human food and in genetic dissection of the great oil trait.

XX SQ. Sequence 2001 BP; 546 A; 406 C; 494 G; 554 T; 1 other;

## Query Match

Best Local Similarity 73.6%; Score 16.2; DB 21; Length 2001;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gttggatcacagaattgata 22

||||| ||||| |||||

Db 42 gttggatcacagaattgta 62

## RESULT 11

T64683  
 ID T64683 standard; DNA; 3822 BP.

XX AC T64683;

XX DT 17-MAR-1998 (first entry)

XX M. leprae gyra precursor coding sequence.

DE XX Mycobacterium sp.; internal sequence; intein; immature; gyrase;

KW protein splicing; precursor; gyra; ss.

KW Mycobacterium leprae.

XX OS FR2739859-A1.

XX PN 18-APR-1997.

XX PD 17-OCT-1995; 95FR-0012162.

XX PF 17-OCT-1995; 95FR-0012162.

XX PR 17-OCT-1995; 95FR-0012162.

XX PA (INSP ) INST PASTEUR.

XX PI Cole S, Fsihi H;

XX DR WPI; 1997-247976/23.

XX DR P-PSDB; W15078.

XX PT Mycobacterial DNA gyrase precursor protein - and polypeptide(s)

PT corresponding to mycobacterial DNA gyrase intein sequences

XX PS Claim 5; Pages 33-34; 55pp; French.

XX CC This DNA, isolated from Mycobacterium leprae, encodes a precursor

CC (immature) gyrase protein from which an "intein" (see W15074) is excised during maturation. The gyrase and its coding sequence are useful to fabricate products that alter the maturation of proteins essential for the development of infectious agents by altering their protein splicing of precursor polypeptides of the proteins. The inteins encoded by the gyra genes of Mycobacterium leprae, M. flavesens, M. goodii and M. kansasii are used: (a) to modify the genome of a eukaryotic cell that lacks endogenous biological activity identical to that of the polypeptide; (b) to replace a copy of a gene present in a recipient genome by integration of a gene different from that where the integration takes place; and (c) for targetted insertion of a foreign DNA sequence into a selected site in the genome of a eukaryotic cell not containing the specific cleavage sites of the polypeptide.

XX SQ Sequence 3822 BP; 843 A; 899 C; 1092 G; 988 T; 0 other;

## Query Match

Best Local Similarity 73.6%; Score 16.2; DB 18; Length 3822;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gttggatcacagaattgata 22

||||| ||||| |||||

Db 2334 gttggatcacagaattgta 2354

## RESULT 12

Q42933  
 ID Q42933 standard; DNA; 4346 BP.

XX AC Q42933;

XX DT 15-OCT-1993 (first entry)

XX DE A3 maize ACCase cDNA clone.

Wed Mar 28 14:03:15 2001

XX Acetyl CoA carboxylase; ACCase; probe: expression; oil content;  
 KW lower; modified; increased; oilseed rape; sunflower; resistant;  
 KW resistance; grass-weed herbicides; fluzafop; alkylketones; maize;  
 KW wheat; barley; ss.  
 XX

XX Zea mays.

XX Location/Qualifiers

XX Key 3..3947

XX CDS /\*tag= a

XX /\*note= "ACCase"

XX misc\_feature 1875..1878

XX /\*tag= b

XX /\*note= "GAN encodes Glu or Asp"

XX WO9311243-A.

XX

XX 10-JUN-1993.

XX

XX 27-NOV-1992; 92WO-GB02205.

XX

XX 28-NOV-1991; 91GB-0025330.

XX

XX (ICIL ) ICI AUSTRALIA OPERATIONS PTY LTD.

XX (ICIL ) IMPERIAL CHEM IND PLC.

XX

XX Ashton AR, Jenkins CL, Whitfield PR;

XX

XX WPI; 1993-197061/24.

XX P-PSDB; R36781.

XX

XX DNA clones for use in probing plant DNA - comprise maize acetyl

XX coenzyme-A carboxylase gene flanked by heterologous DNA

XX Claim 1; Fig 3; 69pp; English.

XX The sequence is that of the A3 acetyl CoA carboxylase (ACCase) cDNA  
 CC maize clone which can be used to probe plant DNA to isolate other  
 CC such clones. It may be used with plant regulatory sequences to  
 CC produce expression cassettes for ACCase. These cassettes are used to  
 CC transform plants to downregulate prodn. of ACCase. This can alter  
 CC the composition of seeds or other plant parts, e.g. enabling oil-  
 CC bearing plants such as oilseed rape, sunflower or oilpalm, having a  
 CC lower or modified oil content to be produced. It can be used to form  
 CC expression cassettes for overexpression of ACCase., leading to prodn.  
 CC of plants with an increased oil content. It may also be used to recover  
 CC the ACCase gene promoter. This can be used to generate RNA in a  
 CC tissue-specific or developmentally regulated manner, and this RNA  
 CC used to inhibit ACCase expression. Monocotyledonous plants can be made  
 CC resistant to grass-weed herbicides such as fluzafop and alkylketones  
 CC by transforming with cassettes adapted to express ACCase. This is  
 CC achieved by overexpression of monocot ACCase, expression of dicot  
 CC ACCase (which is relatively tolerant to these herbicides), or  
 CC possibly by expression of a resistant form of maize ACCase. This  
 CC provides crop plants which are resistant to herbicides used to  
 CC destroy weeds growing between the plants. This allows the overall  
 CC applicn. of herbicide without affecting the crop, and is also useful  
 CC where there has been short term carryover of herbicide from the  
 CC previous crop. Using the method, weeds such as wild oats may be  
 CC controlled, and the farmer is given extra options to obtain an improved  
 CC harvest using means which are safer, cheaper or more effective.  
 CC Suitable crops are maize, wheat and barley.

XX Sequence 4346 BP; 1239 A; 829 C; 1064 G; 1213 T; 1 other;

XX Query Match 73.6%; Score 16.2; DB 14; Length 4346;

XX Best Local Similarity 85.7%; Pred. NO. 55;

XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX 2 gttggatcacagaagttgtta 22

||||| ||||| ||| ||

Db 902 gttggatcacagaagttgtta 922

RESULT 13

T39905

ID T39905 standard; cDNA; 7470 BP.

XX

XX T39905;

XX

XX 23-JAN-1997 (first entry)

XX

XX Maize acetyl CoA carboxylase cDNA.

XX Acetyl CoA carboxylase; ACCase; herbicide tolerance;

XX cyclohexanedione; aryloxyphenoxypionic acid; vegetable oil;

XX oilseed; maize; corn; ss.

XX

XX Zea mays line A188.

XX

XX Key Location/Qualifiers

XX FH 37..7014

XX CDS /\*tag= a

XX /\*EC\_number= 6.4.1.2

XX

XX WO9631609-A2.

XX

XX 10-OCT-1996.

XX

XX 04-APR-1996; 96WO-US04625.

XX

XX 05-APR-1995; 95US-0417089.

XX

XX (MINU ) UNIV MINNESOTA.

XX

XX Egli MA, Gengenbach BG, Gronwald JW, Lutz SM, Somers DA;

XX WPI; 1996-465030/46.

XX P-PSDB; W05590.

XX

XX DNA encoding maize acetyl coenzyme A carboxylase gene - used for  
 prodn. of plants with herbicide tolerance or altered oil content  
 Claim 2; Page 78-80; 131pp; English.

XX A cDNA sequence (T39905) codes for maize acetyl CoA carboxylase  
 (ACCase) (W05590), an enzyme that plays a central role in fatty  
 acid biosynthesis and accumulation in plants and seeds. The  
 complete sequence was deduced from cDNA clone #18-5, obt'd. from  
 a lambda gt11 cDNA library of maize inbred A188 seedlings, and from  
 PCR amplifications (see also T39918-23). Genomic ACCase DNA  
 fragments have also been isolated (see also T39906-16). ACCase DNA  
 can be incorporated into a vector and used to increase the  
 herbicide tolerance or oil content of a transgenic plant, or used  
 to produce ACCase in heterologous systems.

XX Sequence 7470 BP; 2119 A; 1429 C; 1844 G; 2071 T; 7 other;

XX Query Match 73.6%; Score 16.2; DB 17; Length 7470;

XX Best Local Similarity 85.7%; Pred. NO. 58;

XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX 2 gttggatcacagaagttgtta 22

||||| ||||| ||| ||

Db 3972 gttggatcacagaagttgtta 3992

RESULT 14

V29317

ID V29317 standard; cDNA; 7470 BP.

XX

XX V29317;

XX

```

XX 30-JUL-1998 (first entry)
XX Maize ACCase enzyme encoding cDNA.
XX ACCase; maize; herbicide resistant; corn plant; tolerance; Acc1; Acc2;
XX cyclohexanedione; CHD; aryloxyphenoxypropanoic acid; APA; enzyme; ss.
XX Zea mays.
XX Key Location/Qualifiers
XX CDS 37..7014
XX /*tag= a
XX /transl_except= (pos:4339..4341, aa: His)
XX /transl_except= (pos:4456..4458, aa: Ser)
XX /transl_except= (pos:4359..4461, aa: Lys)
XX /product= "ACCase"
XX
XX W09808963-A1.
XX
XX 05-MAR-1998.
XX
XX 29-AUG-1997; 97WO-US15344.
XX
XX 30-AUG-1996; 96US-0697826.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Egli MA, Gengenbach BG, Lutz SM, Marshall LC, Parker WB;
XX Somers DA, Vandee KL, Wyse DL;
XX
XX WPI; 1998-207043/18.
XX P-PSDB; W56736.
XX
XX Herbicide resistant corn plants - prepared using Acc1 and Acc2 gene
XX combinations
XX
XX Example 7; Fig 3A-E; 112pp; English.
XX
XX This cDNA encodes a maize ACCase enzyme. This can be used in a method
XX of preparing an herbicide resistant corn plant which comprises crossing a
XX first corn plant to a second corn plant so as to yield a progeny plant,
XX where the first corn plant comprises at least 1 herbicide resistant
XX allele and the second plant comprises at least 1 herbicide resistant
XX allele which is not allelic to the herbicide resistant allele present in
XX the first plant. The herbicide resistant corn plants are prepared using
XX Acc1 and Acc2 gene combinations which impart cyclohexanedione (CHD) or
XX aryloxyphenoxypropanoic acid (APA) herbicide tolerance to the corn plant.
XX The methods are used to impart CHD and APA herbicide tolerance to corn
XX plants and to produce CHD or APA herbicide tolerant Zea mays (corn)
XX homozygous or heterozygous for Acc1 and homozygous or heterozygous for
XX Acc2. The methods can also be used to impart tolerance to a corn plant
XX to an agent which inhibits acetyl CoA carboxylase, selected from
XX 3-(2,4-dichlorophenyl)-perhydroindolizine-2,4-dione, 3-isopropyl-6-
XX (N-[2,2-dimethylpropyl]-acetamido-1,3,5-triazine-2,4 (1H,3H)dione,
XX soraphen A and their structural analogues.
XX
XX Sequence 7470 BP; 2119 A; 1429 C; 1844 G; 2071 T; 7 other;
XX
XX Query Match 73.6%; Score 16.2; DB 19; Length 7470;
XX Best Local Similarity 85.7%; Pred. No. 58;
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 2 gttggatcacagaattgata 22
XX : ||||| ||||| ||| ||
XX Db 3972 gttggatcacagaattgatta 3992
XX
XX RESULT 15
XX 249820
XX ID Z49820 standard; cDNA; 7470 BP.
XX

```

```

AC Z49820;
XX
XX 18-APR-2000 (first entry)
XX Maize acetyl CoA carboxylase cDNA.
XX
XX Herbicide resistance; gene modification;
XX maize acetyl CoA carboxylase; ACCase; herbicide; plant oil content;
XX marker-assisted plant selection; groat oil trait;
XX restriction fragment length polymorphism mapping;
XX high-energy animal feed; low-fat human food; ss.
XX Zea mays.
XX
XX Key Location/Qualifiers
XX CDS 37..7014
XX /*tag= a
XX /product= "Maize acetyl CoA carboxylase"
XX 1..36
XX /*tag= b
XX 7014..7470
XX /*tag= c
XX
XX 5'UTR
XX
XX 3'UTR
XX
XX W09967367-A1.
XX
XX 29-DEC-1999.
XX
XX 22-JUN-1999; 99WO-US14022.
XX
XX 22-JUN-1998; 98US-0090240.
XX 02-JUL-1998; 98US-0091640.
XX
XX (MINU ) UNIV MINNESOTA.
XX (USDA ) US DEPT OF AGRICULTURE.
XX (EGLI/) EGLI M A.
XX (GROH/) GROH S.
XX (KIAN/) KIANIAN S F.
XX (PHIL/) PHILLIPS R L.
XX (RINE/) RINES H W.
XX (SOME/) SOMERS D A.
XX
XX Egli MA, Groh S, Kianian SF, Phillips RL, Rines HW, Somers DA;
XX
XX WPI; 2000-147205/13.
XX P-PSDB; Y44687.
XX
XX New DNA encoding acetyl-CoA carboxylase from oats, used to produce
XX transformed plants with herbicide resistance and altered oil content
XX
XX Example 6; Fig 13; 197pp; English.
XX
XX The present cDNA sequence encodes maize acetyl CoA carboxylase (ACCase).
XX ACCase is involved in fatty acid synthesis and is the target of
XX the specified herbicides. Transformation of plants with ACCase imparts
XX resistance to cyclohexanedione and aryloxyphenoxypropanoic acid
XX herbicides and alter the oil content. The ACCase cDNA is also used as
XX source of probes and primers for the identification of transgenic
XX plants; in marker-assisted plant selection and for restriction fragment
XX length polymorphism mapping, used for high-energy animal feed and
XX high-fiber, low-fat human food and in genetic dissection of the groat
XX oil trait.
XX
XX Sequence 7470 BP; 2119 A; 1430 C; 1843 G; 2071 T; 7 other;
XX
XX Query Match 73.6%; Score 16.2; DB 21; Length 7470;
XX Best Local Similarity 85.7%; Pred. No. 58;
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 2 gttggatcacagaattgata 22
XX : ||||| ||||| ||| ||
XX Db 3972 gttggatcacagaattgatta 3992
XX

```

us-09-380-826a-2.rng

Wed Mar 28 14:03:15 2001

Search completed: March 27, 2001, 08:22:58  
Job Time: 5300 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:18 ; Search time 132.3 Seconds  
(without alignments)  
26.799 Million cell updates/sec

Title: US-09-380-826A-2

Perfect score: 22

Sequence: 1 ttttgatcacaaagtattgata 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.2	73.6	2000	1	US-08-014-326-1
2	16.2	73.6	2001	3	US-08-417-089-1
3	16.2	73.6	2001	3	US-08-695-651-1
4	16.2	73.6	4345	2	US-08-244-537-1
5	16.2	73.6	7470	3	US-08-417-089-5
6	16.2	73.6	7470	3	US-08-695-651-5
7	15.6	70.9	296	3	US-08-602-145-14
8	15.6	70.9	1096	1	US-08-684-862-8
9	15.6	70.9	1989	2	US-08-792-055-1
10	15.6	70.9	7745	1	US-08-299-675-1
11	15.6	70.9	7745	1	US-08-485-241-1
12	15.6	70.9	7745	2	US-08-874-162-1
13	15.2	69.1	1016	1	US-08-399-986B-3
14	15.2	69.1	1016	1	US-08-493-754A-3
15	15.2	69.1	2182	1	US-08-399-986B-1
16	15.2	69.1	2182	1	US-08-493-754A-1
17	15.2	69.1	2598	3	US-08-745-892-20
18	15.2	69.1	5420	5	5256642-3
19	15.2	69.1	5420	5	5472939-3
20	15.2	69.1	6951	5	5256642-1
21	15.2	69.1	6951	5	5472939-1
22	14.8	67.3	1855	3	US-08-961-083-71
23	14.8	67.3	3159	1	US-08-119-361-4
24	14.8	67.3	3159	3	US-08-336-308A-3
25	14.8	67.3	3159	3	US-08-822-324-3
26	14.8	67.3	7266	3	US-08-336-308A-9
27	14.8	67.3	7266	3	US-08-822-324-5
28	14.8	67.3	8640	1	US-08-570-311-28

29	14.8	67.3	9919	3	US-08-880-179-1	Sequence 1, Appli
30	14.6	66.4	293	3	US-08-866-340-13	Sequence 13, Appli
31	14.6	66.4	1404	1	US-08-204-656B-1	Sequence 1, Appli
32	14.6	66.4	1404	1	US-08-204-656B-3	Sequence 3, Appli
33	14.6	66.4	1404	1	US-08-204-656B-5	Sequence 5, Appli
34	14.6	66.4	1404	1	US-08-204-656B-7	Sequence 7, Appli
35	14.6	66.4	1404	1	US-08-470-702-1	Sequence 1, Appli
36	14.6	66.4	1404	1	US-08-470-702-2	Sequence 2, Appli
37	14.6	66.4	1404	1	US-08-470-702-3	Sequence 3, Appli
38	14.6	66.4	1404	1	US-08-470-702-4	Sequence 4, Appli
39	14.6	66.4	1404	1	US-08-467-831-1	Sequence 1, Appli
40	14.6	66.4	1404	1	US-08-467-831-2	Sequence 2, Appli
41	14.6	66.4	1404	1	US-08-467-831-3	Sequence 3, Appli
42	14.6	66.4	1404	1	US-08-467-831-4	Sequence 4, Appli
C 43	14.6	66.4	1608	2	US-08-424-224-1	Sequence 1, Appli
C 44	14.6	66.4	1608	4	PCT-US94-02891-68	Sequence 68, Appli
45	14.6	66.4	12047	2	US-09-022-461-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-014-326-1  
; Sequence 1, Application US/08014326  
; Patent No. 5498544  
; GENERAL INFORMATION:  
; APPLICANT: Gengenbach, Burle G.  
; APPLICANT: Somers, David A.  
; APPLICANT: Wyse, Donald L.  
; APPLICANT: Gronwald, John W.  
; APPLICANT: Egli, Margaret A.  
; APPLICANT: Lutz, Shiela M.  
; TITLE OF INVENTION: Method and An Acetyl CoA Carboxylase  
; TITLE OF INVENTION: Alteration in Oil Content of Plants  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 3100 No. 5498544west Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/014,326  
; FILING DATE: 05-FEB-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/917,462  
; FILING DATE: 21-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/538,674  
; FILING DATE: 18-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/269,584  
; FILING DATE: 10-NOV-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kowalchuk, Katherine M.  
; REGISTRATION NUMBER: 36,848  
; REFERENCE/DOCKET NUMBER: 600.258-US01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-332-5300  
; TELEFAX: 612-332-9081  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2000 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

```
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: 2 kb fragment of lambda clone #15-14
; US-08-014-326-1

Query Match      73.6%; Score 16.2; DB 1; Length 2000;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggtacacaagatttgata 22
   ||||| ||||| ||||| |||||
Db 42 GTTGATGACAAAGAGTTGTGA 62

RESULT 2
US-08-417-089-1
; Sequence 1, Application US/08417089
; Patent No. 6089298
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN
; TITLE OF INVENTION: OIL CONTENT OF PLANTS
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-417-089-1

Query Match      73.6%; Score 16.2; DB 3; Length 2001;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggtacacaagatttgata 22
   ||||| ||||| ||||| |||||
Db 42 GTTGATGACAAAGAGTTGTGA 62

RESULT 3
US-08-695-651-1
; Sequence 1, Application US/08695651
; Patent No. 6146867
; GENERAL INFORMATION:
; APPLICANT: Gengenbach, B. G.
; APPLICANT: Somers, D. A.
; APPLICANT: Wyse, D. L.
; APPLICANT: Gronwald, J. W.
; APPLICANT: Egli, M. A.
; APPLICANT: Lutz, S. M.
; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/695,651
; FILING DATE: 12-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/417089
; FILING DATE: 05-APR-1995
; APPLICATION NUMBER: 08/014326
; FILING DATE: 05-FEB-1993
; APPLICATION NUMBER: 07/917462
; FILING DATE: 21-JUL-1992
; APPLICATION NUMBER: 07/538674
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.318US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-373-6900
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-695-651-1

Query Match      73.6%; Score 16.2; DB 3; Length 2001;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggtacacaagatttgata 22
   ||||| ||||| ||||| |||||
Db 42 GTTGATGACAAAGAGTTGTGA 62

RESULT 4
US-08-244-537-1
; Sequence 1, Application US/08244537
; Patent No. 5854420
; GENERAL INFORMATION:
; APPLICANT: ASHTON, ANTHONY R.
; APPLICANT: JENKINS, COLIN L.D.
; APPLICANT: WHITFIELD, PAUL R.
; TITLE OF INVENTION: MAIZE ACETYL COA CARBOXYLASE ENCODING
; TITLE OF INVENTION: DNA CLONES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,537
; FILING DATE: 18-AUG-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9125330.2
```



;; FILING DATE: 28-NOV-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/GB92/02205  
;; FILING DATE: 27-NOV-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: KOKULIS, PAUL N.  
;; REGISTRATION NUMBER: 16,773  
;; REFERENCE/DOCKET NUMBER: 203094/SEE 36663/UST  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 861-3000  
;; TELEFAX: (202) 822-0944  
;; TELEX: 6714627 CUSH  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4345 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
US-08-244-537-1

Query Match 73.6%; Score 16.2; DB 2; Length 4345;  
Best Local Similarity 85.7%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22  
||||| ||||| ||| ||  
Db 902 GTTGGATGACAAGAGTTGTTA 922

RESULT 5  
US-08-417-089-5  
; Sequence 5, Application US/08417089  
; Patent No. 6069298  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE  
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN  
; TITLE OF INVENTION: OIL CONTENT OF PLANTS  
; NUMBER OF SEQUENCES: 11  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/417,089  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7470 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-417-089-5

Query Match 73.6%; Score 16.2; DB 3; Length 7470;  
Best Local Similarity 85.7%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22  
||||| ||||| ||| ||  
Db 3972 GTTGGATGACAAGAGTTGTTA 3992

RESULT 6  
US-08-695-651-5  
; Sequence 5, Application US/08695651  
; Patent No. 6146867  
; GENERAL INFORMATION:  
; APPLICANT: Gengenbach, B. G.

;; APPLICANT: Somers, D. A.  
;; APPLICANT: Wyse, D. L.  
;; APPLICANT: Gronwald, J. W.  
;; APPLICANT: Egli, M. A.  
;; APPLICANT: Lutz, S. M.  
;; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA  
;; NUMBER OF SEQUENCES: 11  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.  
;; STREET: P.O. Box 2938  
;; CITY: Minneapolis  
;; STATE: MN  
;; COUNTRY: USA  
;; ZIP: 55402  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSEQ Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/695,651  
;; FILING DATE: 12-AUG-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/417089  
;; FILING DATE: 05-APR-1995  
;; APPLICATION NUMBER: 08/014326  
;; FILING DATE: 05-FEB-1993  
;; APPLICATION NUMBER: 07/917462  
;; FILING DATE: 21-JUL-1992  
;; APPLICATION NUMBER: 07/538674  
;; FILING DATE: 18-JUN-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Woessner, Warren D  
;; REGISTRATION NUMBER: 30,440  
;; REFERENCE/DOCKET NUMBER: 600.318US3  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 612-373-6900  
;; TELEFAX: 612-339-3061  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 7470 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: CDNA  
US-08-695-651-5

Query Match 73.6%; Score 16.2; DB 3; Length 7470;  
Best Local Similarity 85.7%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22  
||||| ||||| ||| ||  
Db 3972 GTTGGATGACAAGAGTTGTTA 3992

RESULT 7  
US-08-602-145-14/c  
; Sequence 14, Application US/08602145  
; Patent No. 6025336  
; GENERAL INFORMATION:  
; APPLICANT: Goltry, Kristin L.  
; APPLICANT: Greenberger, Joel S.  
; TITLE OF INVENTION: DETERMINING EXPOSURE TO IONIZING RADIATION  
; TITLE OF INVENTION: AGENT WITH PERSISTENT BIOLOGICAL MARKERS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington



Query Match 70.9%; Score 15.6; DB 2; Length 1989;  
Best Local Similarity 81.8%; Pred. No. 39;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tggtagatcacaaagatttgata 22  
||||| ||||| ||||| ||  
Db 288 TGTGGATCAACAGATTGTA 267

RESULT 10  
US-08-299-675-1/c  
; Sequence 1, Application US/08299675  
; Patent No. 5389618  
; GENERAL INFORMATION:  
; APPLICANT: Hannah, L. Curtis  
; TITLE OF INVENTION: Materials and Methods for Increasing  
; TITLE OF INVENTION: Corn Seed Weight  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/299,675  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: UF146  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7745 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-299-675-1

Query Match 70.9%; Score 15.6; DB 1; Length 7745;  
Best Local Similarity 81.8%; Pred. No. 46;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tggtagatcacaaagatttgata 22  
||||| ||||| ||||| ||  
Db 3600 TGTGGATAACAAGATGACATA 3579

RESULT 11  
US-08-485-241-1/c  
; Sequence 1, Application US/08485241  
; Patent No. 5650557  
; GENERAL INFORMATION:  
; APPLICANT: Hannah, L. Curtis  
; TITLE OF INVENTION: Materials and Methods for Increasing  
; TITLE OF INVENTION: Corn Seed Weight  
; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/485,241  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/299,675  
; FILING DATE: 1-SEP-1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pace, Doran R.  
; REGISTRATION NUMBER: 38,261  
; REFERENCE/DOCKET NUMBER: UF146.C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7745 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-485-241-1

Query Match 70.9%; Score 15.6; DB 1; Length 7745;  
Best Local Similarity 81.8%; Pred. No. 46;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tggtagatcacaaagatttgata 22  
||||| ||||| ||||| ||  
Db 3600 TGTGGATAACAAGATGACATA 3579

RESULT 12  
US-08-874-162-1/c  
; Sequence 1, Application US/08874162  
; Patent No. 5872216  
; GENERAL INFORMATION:  
; APPLICANT: Hannah, L. Curtis  
; APPLICANT: Giroux, Michael  
; TITLE OF INVENTION: Materials and Methods for Increasing  
; TITLE OF INVENTION: Corn Seed Weight  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/874,162  
; FILING DATE: 13-JUN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/485,241  
 FILING DATE: 7-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/299,675  
 FILING DATE: 1-SEP-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Pace, Doran K.  
 REGISTRATION NUMBER: 38,261  
 REFERENCE/DOCKET NUMBER: UF-146C1D1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 352-375-8100  
 TELEFAX: 352-372-5800  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7745 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-874-162-1

Query Match 70.9%; Score 15.6; DB 2; Length 7745;  
 Best Local Similarity 81.8%; Pred. No. 46;  
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagatttgata 22  
 ||||| ||||| ||||| |||||  
 DB 3600 TTTTGGATAACAAGATGACATA 3579

RESULT 13  
 US-08-399-986B-3/C  
 ; Sequence 3, Application US/08399986B  
 ; Patent No. 5801041  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godwin, Andrew K.  
 ; TITLE OF INVENTION: No. 5801041el Gene Associated with Suppression  
 ; TITLE OF INVENTION: of Tumor Development  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
 ; STREET: 1601 Market Street  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103-2307  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/399,986B  
 ; FILING DATE: 06-MAR-1995  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hagan, Patrick J.  
 ; REGISTRATION NUMBER: 27,643  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 563-4100  
 ; TELEFAX: (215) 563-4044  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1016 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO

US-08-399-986B-3

Query Match 69.1%; Score 15.2; DB 1; Length 1016;  
 Best Local Similarity 85.0%; Pred. No. 57;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttggatcacaaagatttgata 22  
 ||||| ||||| ||||| |||||  
 DB 830 TTGGTCTCAAGAAATTGATA 811

RESULT 14  
 US-08-493-754A-3/C  
 ; Sequence 3, Application US/08493754A  
 ; Patent No. 5821338  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godwin, Andrew K.  
 ; TITLE OF INVENTION: No. 5821338el Gene Associated with Suppression  
 ; TITLE OF INVENTION: of Tumor Development  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
 ; STREET: 1601 Market Street  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103-2307  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/493,754A  
 ; FILING DATE: 22-JUN-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hagan, Patrick J.  
 ; REGISTRATION NUMBER: 27,643  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 563-4100  
 ; TELEFAX: (215) 563-4044  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1016 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; US-08-493-754A-3

Query Match 69.1%; Score 15.2; DB 1; Length 1016;  
 Best Local Similarity 85.0%; Pred. No. 57;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttggatcacaaagatttgata 22  
 ||||| ||||| ||||| |||||  
 DB 830 TTGGTCTCAAGAAATTGATA 811

RESULT 15  
 US-08-399-986B-1/c  
 ; Sequence 1, Application US/08399986B  
 ; Patent No. 5801041  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godwin, Andrew K.  
 ; TITLE OF INVENTION: No. 5801041el Gene Associated with Suppression  
 ; TITLE OF INVENTION: of Tumor Development  
 ; NUMBER OF SEQUENCES: 35

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,986B
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2182 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-399-986B-1

```

```

Query Match      69.1%; Score 15.2; DB 1; Length 2182;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy   3 ttggatcacagatttgata 22
      ||||| || ||||| |||||
Db   1996 TTGGGCTCTCAAGATTGATA 1977

```

Search completed: March 27, 2001, 08:19:19  
Job time: 5146 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:19 ; Search time 132.3 Seconds  
(without alignments)  
8.527 Million cell updates/sec

Title: US-09-380-826A-4  
Perfect score: 7 tgttggg 7  
Sequence: 1 tgttggg 7  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	7	100.0	10	2	US-08-631-751A-13
C 2	7	100.0	14	1	US-08-303-004-19
C 3	7	100.0	14	3	US-08-765-340-154
C 4	7	100.0	15	1	US-08-182-968A-201
C 5	7	100.0	15	1	US-08-182-968A-246
C 6	7	100.0	15	1	US-08-363-240A-614
C 7	7	100.0	15	1	US-08-363-240A-615
C 8	7	100.0	15	1	US-08-363-240A-616
C 9	7	100.0	15	2	US-08-774-306A-201
C 10	7	100.0	15	2	US-08-774-306A-246
C 11	7	100.0	15	2	US-08-585-684B-1366
C 12	7	100.0	15	2	US-08-585-684B-1367
C 13	7	100.0	15	2	US-08-585-684B-1729
C 14	7	100.0	15	2	US-08-585-684B-1730
C 15	7	100.0	15	2	US-08-585-684B-1731
C 16	7	100.0	15	2	US-08-585-684B-1732
C 17	7	100.0	15	2	US-08-585-684B-1768
C 18	7	100.0	15	2	US-08-585-684B-1769
C 19	7	100.0	15	2	US-08-585-684B-1770
C 20	7	100.0	15	2	US-08-585-684B-2139
C 21	7	100.0	15	3	US-09-064-156A-201
C 22	7	100.0	15	3	US-09-064-156A-246
C 23	7	100.0	16	1	US-08-435-350-32
C 24	7	100.0	16	1	US-08-166-664-15
C 25	7	100.0	16	2	US-08-469-461-15
C 26	7	100.0	16	3	US-07-890-609-15
C 27	7	100.0	17	1	US-08-072-063-4
C 28	7	100.0	17	1	US-08-119-773-23

C 29	7	100.0	17	1	US-08-281-940-25	Sequence 25, Appl
C 30	7	100.0	17	1	US-08-390-850-433	Sequence 433, App
C 31	7	100.0	17	1	US-08-064-693-4	Sequence 4, Appli
C 32	7	100.0	17	1	US-08-373-124A-810	Sequence 810, App
C 33	7	100.0	17	1	US-08-441-370-2	Sequence 2, Appli
C 34	7	100.0	17	1	US-08-435-634-433	Sequence 433, App
C 35	7	100.0	17	1	US-08-435-628-810	Sequence 810, App
C 36	7	100.0	17	2	US-08-710-134-25	Sequence 25, Appl
C 37	7	100.0	17	2	US-08-485-885-25	Sequence 25, Appl
C 38	7	100.0	17	3	US-08-704-966-5	Sequence 5, Appli
C 39	7	100.0	17	3	US-08-705-438-5	Sequence 5, Appli
C 40	7	100.0	17	3	US-08-985-162-529	Sequence 529, App
C 41	7	100.0	17	3	US-08-985-162-530	Sequence 530, App
C 42	7	100.0	17	3	US-08-985-162-531	Sequence 531, App
C 43	7	100.0	17	4	PCT-US93-04754-4	Sequence 4, Appli
C 44	7	100.0	18	1	US-08-361-920-47	Sequence 47, Appl
C 45	7	100.0	18	1	US-08-479-939-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1  
US-08-631-751A-13/C  
; Sequence 13, Application US/08631751A  
; Patent No. 5843767  
; GENERAL INFORMATION:  
; APPLICANT: Beattie, Kenneth L.  
; TITLE OF INVENTION: Microfabricated, Flowthrough Porous  
; TITLE OF INVENTION: Apparatus for Discrete Detection of Binding Reactions  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Vinson & Elkins  
; STREET: 1455 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004-1008  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/631,751A  
; FILING DATE: 11-April-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sanzo, Michael A.  
; REGISTRATION NUMBER: 36,912  
; REFERENCE/DOCKET NUMBER: HARC0001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)639-6500  
; TELEFAX: (202)639-6604  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
US-08-631-751A-13

Query Match 100.0%; Score 7; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 5.1e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tgttggg 7  
Db 7 TGTGTGA 1  
RESULT 2

US-08-303-004-19/C  
; Sequence 19, Application US/08303004  
; Patent No. 5556955  
; GENERAL INFORMATION:  
; APPLICANT: Vergnaud, Gilles  
; TITLE OF INVENTION: Process for Detection of New Polymor-  
; TITLE OF INVENTION: Phic Loci in an ADN Sequence, Nucleotide Sequences Forming  
; TITLE OF INVENTION: Hybridisation Probes and Their Biological Applications  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Oliff & Berridge  
; STREET: P.O. Box 19928  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; APPLICATION DATA:  
; FILING DATE: US/08/303,004  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/931,311B  
; FILING DATE: 19920818  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berridge, William P.  
; REGISTRATION NUMBER: 30,024  
; REFERENCE/DOCKET NUMBER: WPB 28264  
; TELEPHONE: (703) 836-6400  
; TELEFAX: (703) 836-2787  
; TELEX: 90-1799 PTO ALEX  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-303-004-19

Query Match 100.0%; Score 7; DB 1; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5.2e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7  
Db 12 TGTGGA 6

RESULT 3  
US-08-765-340-154  
; Sequence 154, Application US/08765340  
; Patent No. 6150092  
; GENERAL INFORMATION:  
; APPLICANT: UCHIDA, K.,  
; APPLICANT: UCHIDA, T.,  
; APPLICANT: TANAKA, Y.,  
; APPLICANT: MATSUDA, Y.,  
; APPLICANT: KONDO, S.,  
; TITLE OF INVENTION: AN ANTISENSE NUCLEIC ACID  
; NUMBER OF SEQUENCES: 185  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
; STREET: 345 PARK AVENUE

US-08-182-968A-201  
; Sequence 201, Application US/08182968A  
; Patent No. 5610054  
; GENERAL INFORMATION:  
; APPLICANT: Draper, Kenneth G.  
; TITLE OF INVENTION: METHOD AND REAGENT FOR  
; TITLE OF INVENTION: INHIBITING HEPATITIS C  
; TITLE OF INVENTION: VIRUS REPLICATION  
; NUMBER OF SEQUENCES: 497  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,968A  
; FILING DATE: 13-JANUARY-1994

CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version  
SOFTWARE: #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,340  
FILING DATE: 23-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 145146/94  
FILING DATE: 27-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 311130/94  
FILING DATE: 21-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: SERUNIAN, LESLIE  
REGISTRATION NUMBER: 35,353  
REFERENCE/DOCKET NUMBER: 1452-4005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 154:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "synthetic DNA"  
US-08-765-340-154

Query Match 100.0%; Score 7; DB 3; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5.2e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttga 7  
Db 7 TGTGGA 13

RESULT 4  
US-08-182-968A-201  
; Sequence 201, Application US/08182968A  
; Patent No. 5610054  
; GENERAL INFORMATION:  
; APPLICANT: Draper, Kenneth G.  
; TITLE OF INVENTION: METHOD AND REAGENT FOR  
; TITLE OF INVENTION: INHIBITING HEPATITIS C  
; TITLE OF INVENTION: VIRUS REPLICATION  
; NUMBER OF SEQUENCES: 497  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,968A  
; FILING DATE: 13-JANUARY-1994



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/882,888  
FILING DATE: 14-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 205/277  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 201:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-182-968A-201

Query Match 100.0%; Score 7; DB 1; Length 15;  
Best Local Similarity 57.1%; Pred. No. 5.2e+03;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
DB 5 UGUUGGA 11

RESULT 5  
US-08-182-968A-246/c  
Sequence 246, Application US/08182968A  
Patent No. 5610054  
GENERAL INFORMATION:  
APPLICANT: Draper, Kenneth G.  
TITLE OF INVENTION: METHOD AND REAGENT FOR  
TITLE OF INVENTION: INHIBITING HEPATITIS C  
TITLE OF INVENTION: VIRUS REPLICATION  
NUMBER OF SEQUENCES: 497  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/182,968A  
FILING DATE: 13-JANUARY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/882,888  
FILING DATE: 14-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 205/277  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 246:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-182-968A-246

Query Match 100.0%; Score 7; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.2e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
DB 13 TGTGGGA 7

RESULT 6  
US-08-363-240A-614/c  
Sequence 614, Application US/08363240A  
Patent No. 5705388  
GENERAL INFORMATION:  
APPLICANT: Couture, Larry  
APPLICANT: McSwiggen, James  
APPLICANT: Bisgaller, Charles  
APPLICANT: Pape, Michael  
TITLE OF INVENTION: METHOD AND REAGENT FOR  
TITLE OF INVENTION: PREVENTION, INHIBITION OF  
TITLE OF INVENTION: PROGRESSION AND REGRESSION  
TITLE OF INVENTION: OF VASCULAR DISEASES  
NUMBER OF SEQUENCES: 1243  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: Storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/363,240A  
FILING DATE: December 23, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 210/096  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 614:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-363-240A-614

Query Match 100.0%; Score 7; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.2e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgttggg 7  
DB 15 TGTGGGA 9

RESULT 7

US-08-363-240A-615/C  
 ; Sequence 615, Application US/08363240A  
 ; Patent No. 5705388  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Couture, Larry  
 ; APPLICANT: McSwiggen, James  
 ; APPLICANT: Bisgaier, Charles  
 ; APPLICANT: Pape, Michael  
 ; TITLE OF INVENTION: METHOD AND REAGENT FOR  
 ; TITLE OF INVENTION: PREVENTION, INHIBITION OF  
 ; TITLE OF INVENTION: PROGRESSION AND REGRESSION  
 ; TITLE OF INVENTION: OF VASCULAR DISEASES  
 ; NUMBER OF SEQUENCES: 1243  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: 633 West Fifth Street  
 ; CITY: Suite 4700  
 ; STATE: Los Angeles  
 ; COUNTRY: California  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/363,240A  
 ; FILING DATE: December 23, 1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Warburg, Richard  
 ; REGISTRATION NUMBER: 32,327  
 ; REFERENCE/DOCKET NUMBER: 210/096  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (213) 489-1600  
 ; TELEFAX: (213) 955-0440  
 ; TELEX: 67-3510  
 ; INFORMATION FOR SEQ ID NO: 615:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 15 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-363-240A-615

Query Match 100.0%; Score 7; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.2e+03;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 tgttggg 7  
 Db 15 TGTGGG 9  
 RESULT 8  
 US-08-363-240A-616/C  
 ; Sequence 616, Application US/08363240A  
 ; Patent No. 5705388  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Couture, Larry  
 ; APPLICANT: McSwiggen, James  
 ; APPLICANT: Bisgaier, Charles  
 ; APPLICANT: Pape, Michael  
 ; TITLE OF INVENTION: METHOD AND REAGENT FOR  
 ; TITLE OF INVENTION: PREVENTION, INHIBITION OF  
 ; TITLE OF INVENTION: PROGRESSION AND REGRESSION  
 ; TITLE OF INVENTION: OF VASCULAR DISEASES  
 ; NUMBER OF SEQUENCES: 1243

US-08-363-240A-615/C  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: 633 West Fifth Street  
 ; CITY: Suite 4700  
 ; STATE: Los Angeles  
 ; COUNTRY: California  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/363,240A  
 ; FILING DATE: December 23, 1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Warburg, Richard  
 ; REGISTRATION NUMBER: 32,327  
 ; REFERENCE/DOCKET NUMBER: 210/096  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (213) 489-1600  
 ; TELEFAX: (213) 955-0440  
 ; TELEX: 67-3510  
 ; INFORMATION FOR SEQ ID NO: 616:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 15 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-363-240A-616  
 Query Match 100.0%; Score 7; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.2e+03;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 tgttggg 7  
 Db 13 TGTGGG 7  
 RESULT 9  
 US-08-774-306A-201  
 ; Sequence 201, Application US/08774306A  
 ; Patent No. 5869253  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Draper, Kenneth G.  
 ; TITLE OF INVENTION: METHOD AND REAGENT FOR  
 ; TITLE OF INVENTION: INHIBITING HEPATITIS C  
 ; TITLE OF INVENTION: VIRUS REPLICATION  
 ; NUMBER OF SEQUENCES: 497  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: 633 West Fifth Street  
 ; CITY: Suite 4700  
 ; STATE: Los Angeles  
 ; COUNTRY: California  
 ; ZIP: 90071-2066  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/774,306A  
 ; FILING DATE: December 26, 1996

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/182,968
; FILING DATE: January 13, 1994
; APPLICATION NUMBER: 07/882,888
; FILING DATE: May 14, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 223/227
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 201:
; LENGTH: 15
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-774-306A-201

```

```

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 57.1%; Pred. No. 5.2e+03;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tqtgga 7
   :|:|:|
Db 5 UGUUGGA 11

```

```

RESULT 10
US-08-774-306A-246/c
; Sequence 246, Application US/08774306A
; Patent No. 5869253
; GENERAL INFORMATION:
; APPLICANT: Draper, Kenneth G.
; TITLE OF INVENTION: METHOD AND REAGENT FOR
; TITLE OF INVENTION: INHIBITING HEPATITIS C
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/774,306A
; FILING DATE: December 26, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/182,968
; FILING DATE: January 13, 1994
; APPLICATION NUMBER: 07/882,888
; FILING DATE: May 14, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 223/227
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 246:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 15
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-774-306A-246

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tqtgga 7
   :|:|:|:|
Db 13 TGTGGA 7

RESULT 11
US-08-585-684B-1366/c
; Sequence 1366, Application US/08585684B
; Patent No. 5877021
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Daniel T.
; APPLICANT: Jarvis, Thale
; APPLICANT: McSwiggan, James
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 2751
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/585,684B
; FILING DATE: January 16, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/000,951
; FILING DATE: July 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 218/078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1366:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-585-684B-1366

```

```

Query Match 100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tqtgga 7
   :|:|:|:|
Db 14 TGTGGA 8

```

RESULT 12  
US-08-585-684B-1367/C  
; Sequence 1367, Application US/08585684B  
; Patent No. 5877021  
; GENERAL INFORMATION:  
; APPLICANT: Stinchcomb, Daniel T.  
; APPLICANT: Jarvis, Thale  
; APPLICANT: McSwiggen, James  
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE  
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE  
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES  
; NUMBER OF SEQUENCES: 2751  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Suite 4700  
; STATE: Los Angeles  
; COUNTRY: California  
; COUNTRY: U.S.A.  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/585,684B  
; FILING DATE: January 16, 1996  
; PRIOR APPLICATION NUMBER: 60/000,951  
; FILING DATE: July 7, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 218/078  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 1367:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-585-684B-1367

Query Match 100.0%; Score 7; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.2e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 tgttggga 7  
Db 13 TGTGGA 7

RESULT 13  
US-08-585-684B-1729/C  
; Sequence 1729, Application US/08585684B  
; Patent No. 5877021  
; GENERAL INFORMATION:  
; APPLICANT: Stinchcomb, Daniel T.  
; APPLICANT: Jarvis, Thale  
; APPLICANT: McSwiggen, James  
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE  
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE  
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES  
; NUMBER OF SEQUENCES: 2751  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street  
; CITY: Suite 4700  
; STATE: Los Angeles  
; COUNTRY: California  
; COUNTRY: U.S.A.  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/585,684B  
; FILING DATE: January 16, 1996  
; PRIOR APPLICATION NUMBER: 60/000,951  
; FILING DATE: July 7, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 218/078  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 1729:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-585-684B-1729

Query Match 100.0%; Score 7; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.2e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttggga 7  
Db 13 TGTGGA 7

RESULT 14  
US-08-585-684B-1730/C  
; Sequence 1730, Application US/08585684B  
; Patent No. 5877021  
; GENERAL INFORMATION:  
; APPLICANT: Stinchcomb, Daniel T.  
; APPLICANT: Jarvis, Thale  
; APPLICANT: McSwiggen, James  
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE  
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE  
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES  
; NUMBER OF SEQUENCES: 2751  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Suite 4700  
; STATE: Los Angeles  
; COUNTRY: California  
; COUNTRY: U.S.A.  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/585,684B  
; FILING DATE: January 16, 1996

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/000,951
; FILING DATE: July 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 218/078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1730:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-585-684B-1730

```

```

Query Match      100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tgttga 7
Db 13 TGTGGA 7

```

# RESULT 15

```

US-08-585-684B-1731/C
; Sequence 1731, Application US/08585684B
; Patent No. 5877021
; GENERAL INFORMATION:

```

```

; APPLICANT: Stinchcomb, Daniel T.
; APPLICANT: Jarvis, Thale
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: METHOD AND REAGENT FOR THE
; TITLE OF INVENTION: INDUCTION OF GRAFT TOLERANCE
; TITLE OF INVENTION: AND REVERSAL OF IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 2751
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: LYON & LYON
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/585,684B
; FILING DATE: January 16, 1996

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/000,951
; FILING DATE: July 7, 1995
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 218/078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510

```

```

; INFORMATION FOR SEQ ID NO: 1731:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-585-684B-1731

```

```

Query Match      100.0%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tgttga 7
Db 13 TGTGGA 7

```

```

Search completed: March 27, 2001, 08:19:22
Job time: 5149 sec

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:22:59 ; Search time 207.51 Seconds  
(without alignments)  
12.672 Million cell updates/sec

Title: US-09-380-826A-5

Perfect score: 7  
Sequence: 1 ttgata 7

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

```
1: /cgn2_2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	100.0	7	19	Leptospira nucleot
2	7	100.0	10	21	Human dendritic ce
3	7	100.0	10	21	Metastatic breast
4	7	100.0	10	21	Metastatic breast
5	7	100.0	10	21	Metastatic breast
6	7	100.0	10	21	Metastatic breast
7	7	100.0	10	21	HIV-1 protease gen
8	7	100.0	13	21	rb gene antisense
9	7	100.0	14	19	HIV-1 protease gen
10	7	100.0	14	21	Neuroblastoma spec
11	7	100.0	15	14	Q52243
12	7	100.0	15	16	RSV IC hammerhead
13	7	100.0	15	16	RSV IC hammerhead

13	7	100.0	15	16	T56964	RSV IC hammerhead
14	7	100.0	15	16	T57424	RSV N hammerhead r
15	7	100.0	15	16	T57425	RSV N hammerhead r
16	7	100.0	15	16	T58433	Mouse reIA hammerh
17	7	100.0	15	16	T58435	Mouse reIA hammerh
18	7	100.0	15	16	T58441	Mouse reIA hammerh
19	7	100.0	15	17	X56632	Human CD40 hammerh
20	7	100.0	15	17	X56633	Human CD40 hammerh
21	7	100.0	15	17	X56634	Human CD40 hammerh
22	7	100.0	15	17	X56635	Human CD40 hammerh
23	7	100.0	15	17	X56825	Mouse CD40 hammerh
24	7	100.0	15	17	X56826	Mouse CD40 hammerh
25	7	100.0	15	17	X56827	Mouse CD40 hammerh
26	7	100.0	15	17	X64619	Human B7-1 hammerh
27	7	100.0	15	17	X64620	Human B7-1 hammerh
28	7	100.0	15	17	X64621	Human B7-1 hammerh
29	7	100.0	15	17	X64622	Human B7-1 hammerh
30	7	100.0	15	17	X75694	Human flt-1 and KD
31	7	100.0	15	20	X76904	H2-1 Pagl gene dir
32	7	100.0	15	20	X31695	Transcript tag seq
33	7	100.0	15	20	X31205	Tag sequence Of a
34	7	100.0	15	21	X297690	HIV-1 protease gen
35	7	100.0	15	21	X297727	HIV-1 protease gen
36	7	100.0	16	12	Q11395	Probe COD 1111 Spe
37	7	100.0	16	14	Q52019	B-cell mRNA ribozy
38	7	100.0	16	16	Q74138	Basic fibroblast g
39	7	100.0	16	21	X297686	HIV-1 protease gen
40	7	100.0	16	21	X297687	HIV-1 protease gen
41	7	100.0	16	21	X297689	HIV-1 protease gen
42	7	100.0	16	21	X297702	HIV-1 protease gen
43	7	100.0	16	21	X297704	HIV-1 protease gen
44	7	100.0	16	21	X297721	HIV-1 protease gen
45	7	100.0	16	21	X297722	HIV-1 protease gen

#### ALIGNMENTS

RESULT	1
V58899	ID V58899 standard; DNA; 7 BP.
XX	XX
AC	V58899;
XX	XX
DT	20-JAN-1999 (first entry)
XX	XX
DE	Leptospira nucleotide sequence.
XX	XX
KW	Infection; pathogenic Leptospira; protective immunity; therapy;
XX	diagnosis; ss.
OS	Leptospira sp.
XX	XX
PN	WO9840099-A1.
XX	XX
PD	17-SEP-1998.
XX	XX
PF	06-MAR-1998; 98WO-AU00145.
XX	XX
PR	07-MAR-1997; 97AU-0005494.
XX	XX
PA	(AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
PA	(PIGR-) PIG RES & DEV CORP.
PI	Chappel RJ;
XX	XX
DR	WPI; 1998-520791/44.
XX	XX
PT	New isolated pathogenic Leptospira bacterium - useful for, e.g
PT	developing products for conferring protective immunity, and for
XX	prophylactic or therapeutic treatment
PS	Claim 15; Page 72; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the  
CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
CC belongs to serogroup Hurstbridge or serovar Hurstbridge or the species  
CC L. fainei. The LS bacteria can be used for conferring protective  
CC immunity against pathogenic LS bacteria in humans or animals. The  
CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
CC infections. The DNAs and antibodies may also be used for detection and  
CC diagnosis of past or present LS infection.  
XX  
SQ Sequence 7 BP; 2 A; 0 C; 1 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 7;  
Best Local Similarity 100.0%; Pred. No. 5.2e+07;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7  
Db 1 ttgtgata 7  
|||||

RESULT 2  
278577/C  
ID 278577 standard; DNA; 10 BP.  
XX  
AC 278577;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE Human dendritic cell SAGE tag, SEQ ID NO:1005.  
XX  
KW SAGE tag; serial analysis of gene expression; antigen-presenting cell;  
KW APC; monocyte-derived dendritic cell; differential gene expression;  
KW immunostimulatory cofactor; costimulatory factor; CTL;  
KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9965924-A2.  
XX  
PD 23-DEC-1999.  
XX  
PF 18-JUN-1999; 99WO-US13800.  
XX  
PR 19-JUN-1998; 98US-0089833.  
PR 19-JUN-1998; 98US-0089844.  
PR 19-JUN-1998; 98US-0089853.  
PR 19-JUN-1998; 98US-0089878.  
PR 19-JUN-1998; 98US-0089991.  
PR 19-JUN-1998; 98US-0089992.  
PR 19-JUN-1998; 98US-0089993.  
PR 19-JUN-1998; 98US-0089994.  
PR 19-JUN-1998; 98US-0089997.  
PR 19-JUN-1998; 98US-0089999.  
PR 19-JUN-1998; 98US-0090000.  
PR 19-JUN-1998; 98US-0090035.  
PR 19-JUN-1998; 98US-0090036.  
PR 19-JUN-1998; 98US-0090039.  
PR 19-JUN-1998; 98US-0090040.  
PR 19-JUN-1998; 98US-0090041.  
PR 19-JUN-1998; 98US-0090042.  
PR 19-JUN-1998; 98US-0090043.  
PR 19-JUN-1998; 98US-0090044.  
PR 19-JUN-1998; 98US-0090045.  
PR 19-JUN-1998; 98US-0090047.  
PR 19-JUN-1998; 98US-0090048.  
PR 19-JUN-1998; 98US-0090072.  
PR 19-JUN-1998; 98US-0090076.  
PR 19-JUN-1998; 98US-0090077.  
PR 19-JUN-1998; 98US-0090078.  
PR 19-JUN-1998; 98US-0090079.  
PR 19-JUN-1998; 98US-0090080.

PR 08-DEC-1998; 98US-0111715.  
XX (GENZ ) GENZYME CORP.  
PA (ROBE/) ROBERTS B L.  
PA (SHAN/) SHANKARA S.  
XX  
PI Roberts BL, Shankara S;  
XX WPI; 2000-106077/09.  
XX  
XX Isolated polynucleotides differentially expressed in antigen-presenting  
PT cells, useful in gene vaccines against cancer -  
XX  
XX Claim 1; Page 94; 130pp; English.  
XX  
XX Sequences 277573-279709 represent SAGE (serial analysis of gene  
CC expression) tags used to identify mRNA transcripts encoding or  
CC immunostimulatory cofactor proteins which are preferentially or  
CC differentially expressed in monocyte-derived dendritic cells compared  
CC with monocytes. Some of the transcripts correspond to known genes or  
CC ESTs (expressed sequence tags) which were previously unknown to be  
CC preferentially or differentially expressed in dendritic cells, while  
CC other transcripts correspond to novel genes. Antigen-presenting cell  
CC (APC)-associated costimulatory factors play an important role in the  
CC activation of the cytotoxic immune response, particularly against tumour  
CC cells. Tumour antigen presentation via the MHC (major histocompatibility  
CC complex) and subsequent recognition by T-cell receptors is alone  
CC insufficient to activate a robust cytotoxic immune response that can lyse  
CC the tumour cells. Immunostimulatory cofactors also being required for  
CC efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid  
CC sequences identified using the SAGE tags have several potential uses.  
CC They may be used in vaccines to induce an immune response, particularly  
CC against a tumour antigen; to modulate the genotype of an APC; to screen  
CC for agents that modulate expression of differentially expressed genes in  
CC an APC; and as hybridisation probes/amplification primers for the  
CC diagnosis, prognosis and monitoring of diseases related to abnormal  
CC expression of these genes. Detection of the dendritic cell differentially  
CC expressed genes, or of their encoded proteins, can be used to identify  
CC cells as belonging to the monocyte lineage. Cells containing these genes  
CC can be used in active immunotherapy (or to stimulate production of a  
CC population of antigen-specific effector cells) and vectors containing  
CC them are used in gene therapy. Co-administration of tumour antigens and  
CC APC-associated costimulatory factors ensures adequate antigen  
CC presentation to endogenous APCs and upregulates the APCs for the  
CC presentation of co-stimulatory signals, migration to T cell-rich sites,  
CC secretion of T cell growth factors and secretion of chemokines for  
CC recruitment of immune effector cells.  
XX  
SQ Sequence 10 BP; 6 A; 1 C; 1 G; 2 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7  
Db 9 ttgtgata 3  
|||||

RESULT 3  
280864  
ID 280864 standard; DNA; 10 BP.  
XX  
AC 280864;  
XX  
XX 07-APR-2000 (first entry)  
DE Metastatic breast tumour cell upregulated transcript tag #98.  
XX  
KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
KW antimetastatic; vaccine; diagnosis; ss.



XX Homo sapiens.  
 XX WO9965928-A2.  
 XX 23-DEC-1999.  
 XX 18-JUN-1999; 99WO-US13647.  
 XX 19-JUN-1998; 98US-0089853.  
 XX 19-JUN-1998; 98US-0089997.  
 XX 19-JUN-1998; 98US-0090039.  
 XX 19-JUN-1998; 98US-0090040.  
 XX 19-JUN-1998; 98US-0090041.  
 XX (GENZ ) GENZYME CORP.  
 XX (ROBE/) ROBERTS B L.  
 XX (SHAN/) SHANKARA S.  
 XX Roberts BL, Shankara S;  
 XX WPI; 2000-106079/09.  
 XX Isolated polynucleotides differentially expressed between metastatic  
 PT and non-metastatic breast cancer cells, useful for diagnosis,  
 PT prevention and treatment of cancer -  
 XX Claim 1; Page 60; 219pp; English.  
 XX 280767 to 283941 represent tags corresponding to distinct transcripts  
 CC that are preferentially transcribed in the metastatic breast tumour  
 CC tissue (i.e. are upregulated in metastatic breast tumour cells). 283942  
 CC to 286677 represent tags corresponding to distinct transcripts that are  
 CC preferentially transcribed in the primary or non-metastatic breast tumour  
 CC tissue (i.e. are downregulated in metastatic breast tumour cells).  
 CC These transcripts can be used for diagnosis, prognosis, monitoring and  
 CC treatment of breast cancer, particularly where metastatic. Diagnosis is  
 CC by standard immunoassays or hybridisation/amplification reactions.  
 CC Compounds that modulate expression of the transcripts are potentially  
 CC useful for treatment of (metastatic) breast cancer, while promoters from  
 CC the transcripts are used to direct expression, in selected cell types,  
 CC of e.g. therapeutic genes (also ribozymes or antisense sequences),  
 CC particularly an antigen-encoding sequence for use in gene or cell-based  
 CC vaccines. Polypeptides encoded by the transcripts are also useful in  
 CC vaccines; for diagnosing breast cancer and for raising specific  
 CC antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic  
 CC agents. Host cells that produce the polypeptides can be used to expand  
 CC and isolate populations of educated, antigen-specific immune effector  
 CC cells, e.g. cytotoxic T lymphocytes, and these used for adoptive  
 CC immunotherapy.  
 XX Sequence 10 BP; 3 A; 0 C; 1 G; 6 T; 0 other;  
 SQ Query Match 100.0%; Score 7; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred No. 1.9e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ttgtata 7  
 Db 3 ttgtata 9  
 RESULT 4  
 Z82567  
 ID 282567 standard; DNA; 10 BP.  
 AC 282567;  
 XX 07-APR-2000 (first entry)  
 DT Metastatic breast tumour cell upregulated transcript tag #1801.  
 XX

KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
 KW antimetastatic; vaccine; diagnosis; ss.  
 XX Homo sapiens.  
 XX WO9965928-A2.  
 XX 23-DEC-1999.  
 XX 18-JUN-1999; 99WO-US13647.  
 XX 19-JUN-1998; 98US-0089853.  
 XX 19-JUN-1998; 98US-0089997.  
 XX 19-JUN-1998; 98US-0090039.  
 XX 19-JUN-1998; 98US-0090040.  
 XX 19-JUN-1998; 98US-0090041.  
 XX (GENZ ) GENZYME CORP.  
 XX (ROBE/) ROBERTS B L.  
 XX (SHAN/) SHANKARA S.  
 XX Roberts BL, Shankara S;  
 XX WPI; 2000-106079/09.  
 XX Isolated polynucleotides differentially expressed between metastatic  
 PT and non-metastatic breast cancer cells, useful for diagnosis,  
 PT prevention and treatment of cancer -  
 XX Claim 1; Page 107; 219pp; English.  
 XX 280767 to 283941 represent tags corresponding to distinct transcripts  
 CC that are preferentially transcribed in the metastatic breast tumour  
 CC tissue (i.e. are upregulated in metastatic breast tumour cells). 283942  
 CC to 286677 represent tags corresponding to distinct transcripts that are  
 CC preferentially transcribed in the primary or non-metastatic breast tumour  
 CC tissue (i.e. are downregulated in metastatic breast tumour cells).  
 CC These transcripts can be used for diagnosis, prognosis, monitoring and  
 CC treatment of breast cancer, particularly where metastatic. Diagnosis is  
 CC by standard immunoassays or hybridisation/amplification reactions.  
 CC Compounds that modulate expression of the transcripts are potentially  
 CC useful for treatment of (metastatic) breast cancer, while promoters from  
 CC the transcripts are used to direct expression, in selected cell types,  
 CC of e.g. therapeutic genes (also ribozymes or antisense sequences),  
 CC particularly an antigen-encoding sequence for use in gene or cell-based  
 CC vaccines. Polypeptides encoded by the transcripts are also useful in  
 CC vaccines; for diagnosing breast cancer and for raising specific  
 CC antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic  
 CC agents. Host cells that produce the polypeptides can be used to expand  
 CC and isolate populations of educated, antigen-specific immune effector  
 CC cells, e.g. cytotoxic T lymphocytes, and these used for adoptive  
 CC immunotherapy.  
 XX Sequence 10 BP; 3 A; 0 C; 2 G; 5 T; 0 other;  
 SQ Query Match 100.0%; Score 7; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred No. 1.9e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ttgtata 7  
 Db 1 ttgtata 7  
 RESULT 5  
 Z82768  
 ID 282768 standard; DNA; 10 BP.  
 AC 282768;  
 XX 07-APR-2000 (first entry)  
 DT

XX DE Metastatic breast tumour cell upregulated transcript tag #2002.  
 XX XX  
 KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
 KW non-metastatic breast tumour tissue; gene therapy; anticancer;  
 KW antimetastatic; vaccine; diagnosis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9965928-A2.  
 XX  
 PD 23-DEC-1999.  
 XX  
 XX 18-JUN-1999; 99WO-US13647.  
 XX  
 PR 19-JUN-1998; 98US-0089853.  
 PR 19-JUN-1998; 98US-0089997.  
 PR 19-JUN-1998; 98US-0090039.  
 PR 19-JUN-1998; 98US-0090040.  
 PR 19-JUN-1998; 98US-0090041.  
 XX  
 PA (GENZ ) GENZYME CORP.  
 PA (ROBE/) ROBERTS B L.  
 PA (SHAN/) SHANKARA S.  
 XX  
 PI Roberts BL, Shankara S;  
 XX  
 DR WPI; 2000-106079/09.  
 XX  
 XX Isolated polynucleotides differentially expressed between metastatic  
 PT and non-metastatic breast cancer cells, useful for diagnosis,  
 PT prevention and treatment of cancer -  
 XX  
 PS Claim 1: Page 113; 219pp; English.  
 XX  
 CC 280767 to 283941 represent tags corresponding to distinct transcripts  
 CC that are preferentially transcribed in the metastatic breast tumour  
 CC tissue (i.e. are upregulated in metastatic breast tumour cells). 283942  
 CC to 286677 represent tags corresponding to distinct transcripts that are  
 CC preferentially transcribed in the primary or non-metastatic breast tumour  
 CC tissue (i.e. are downregulated in metastatic breast tumour cells).  
 CC These transcripts can be used for diagnosis, prognosis, monitoring and  
 CC treatment of breast cancer, particularly where metastatic. Diagnosis is  
 CC by standard immunoassays or hybridisation/amplification reactions.  
 CC Compounds that modulate expression of the transcripts are potentially  
 CC useful for treatment of (metastatic) breast cancer, while promoters from  
 CC of e.g. therapeutic genes (also ribozymes or antisense sequences),  
 CC particularly an antigen-encoding sequence for use in gene or cell-based  
 CC vaccines. Polypeptides encoded by the transcripts are also useful in  
 CC antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic  
 CC agents. Host cells that produce the polypeptides can be used to expand  
 CC and isolate populations of educated, antigen-specific immune effector  
 CC cells, e.g. cytotoxic T lymphocytes, and these used for adoptive  
 CC immunotherapy.  
 XX  
 SQ Sequence 10 BP; 4 A; 0 C; 2 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgtgata 7

Db 2 ttgtgata 8

RESULT 6

ID 283117  
 XX 283117 standard; DNA; 10 BP.

283117;

07-APR-2000 (first entry)

Metastatic breast tumour cell upregulated transcript tag #2351.

Human; metastatic breast tumour tissue; breast cancer; tag; primer;  
 non-metastatic breast tumour tissue; gene therapy; anticancer;  
 antimetastatic; vaccine; diagnosis; ss.

Homo sapiens.

WO9965928-A2.

23-DEC-1999.

18-JUN-1999; 99WO-US13647.

19-JUN-1998; 98US-0089853.

19-JUN-1998; 98US-0089997.

19-JUN-1998; 98US-0090039.

19-JUN-1998; 98US-0090040.

19-JUN-1998; 98US-0090041.

(GENZ ) GENZYME CORP.

(ROBE/) ROBERTS B L.

(SHAN/) SHANKARA S.

Roberts BL, Shankara S;

WPI; 2000-106079/09.

Isolated polynucleotides differentially expressed between metastatic  
 and non-metastatic breast cancer cells, useful for diagnosis,  
 prevention and treatment of cancer -

Claim 1; Page 122; 219pp; English.

280767 to 283941 represent tags corresponding to distinct transcripts  
 that are preferentially transcribed in the metastatic breast tumour  
 tissue (i.e. are upregulated in metastatic breast tumour cells). 283942  
 to 286677 represent tags corresponding to distinct transcripts that are  
 preferentially transcribed in the primary or non-metastatic breast tumour  
 tissue (i.e. are downregulated in metastatic breast tumour cells).  
 These transcripts can be used for diagnosis, prognosis, monitoring and  
 treatment of breast cancer, particularly where metastatic. Diagnosis is  
 by standard immunoassays or hybridisation/amplification reactions.  
 Compounds that modulate expression of the transcripts are potentially  
 useful for treatment of (metastatic) breast cancer, while promoters from  
 of e.g. therapeutic genes (also ribozymes or antisense sequences),  
 particularly an antigen-encoding sequence for use in gene or cell-based  
 vaccines. Polypeptides encoded by the transcripts are also useful in  
 antibodies (Ab). Ab are used to detect the polypeptides or as therapeutic  
 agents. Host cells that produce the polypeptides can be used to expand  
 and isolate populations of educated, antigen-specific immune effector  
 cells, e.g. cytotoxic T lymphocytes, and these used for adoptive  
 immunotherapy.

Sequence 10 BP; 4 A; 0 C; 1 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgtgata 7

Db 1 ttgtgata 7

RESULT 7

297719/c  
ID 297719 standard; DNA; 13 BP.

XX AC 297719;

XX DT 26-APR-2000 (first entry)

XX DE HIV-1 protease gene probe SEQ ID NO:209.

XX KW Human immunodeficiency virus; HIV; protease; probe; detection;

XX KW drug selected mutation; hybridisation; genotyping; infection;

XX KW drug resistance; ss.

XX OS Human immunodeficiency virus type 1.

XX PN WO9967428-A2.

XX PD 29-DEC-1999.

XX PF 22-JUN-1999; 99WO-EP04317.

XX PR 24-JUN-1998; 98EP-0870143.

XX PA (INNO-) INNOGENETICS NV.

XX PI Stuyver L;

XX DR WPI; 2000-147219/13.

XX PT Detection of drug-selected mutations in the HIV protease gene used to treat HIV infections

XX PS Claim 3; Page 37; 76pp; English.

XX CC The present invention describes the detection of drug-selected mutations in the HIV protease gene. The method of detection allows the simultaneous characterisation of a range of codons involved in drug resistance using sets of probes optimised to function together in a reverse-hybridisation assay. 297517 to 297997 represent specifically claimed probes for use in the assay, and 297479 to 297501 represent specifically claimed HIV protease gene polymorphic nucleotide sequences. 297502 to 297515, and 298004 to 298007, represent PCR primers for the HIV protease gene, and 297516 represents an HIV protease probe used in an example from the present invention. The method, probes and primers can be used for the detection of drug-selected mutations in the HIV protease gene. The method allows the simultaneous characterisation of a range of codons involved in drug resistance. The method may also be used for HIV protease genotyping assays. The probes are able to discriminate between wild type and mutated protease sequences. The method allows rapid and reliable detection of drug-selected mutation in HIV.

XX SQ Sequence 13 BP; 5 A; 1 C; 2 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 13;  
Best Local Similarity 100.0%; Pred. No. 1.8e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttgtgata 7  
|||||||  
DB 12 TTGTGATA 6

RESULT 8  
V49155

ID V49155 standard; DNA; 14 BP.

XX AC V49155;

XX DT 15-OCT-1998 (first entry)

XX DE rb gene antisense oligonucleotide rb-N-103.

XX

KW rb gene; antisense oligonucleotide; modulate; gene expression; ss.

XX Synthetic.

XX OS Homo sapiens.

XX PN EP856579-A1.

XX PD 05-AUG-1998.

XX PF 31-JAN-1997; 97EP-0101531.

XX PR 31-JAN-1997; 97EP-0101531.

XX PA (BIOG-) BIOGOSTIK GES BIOMOLEKULARE DIAGNOSTIK.

XX PI Brysch W, Schlingensiepen K;

XX WPI; 1998-400910/35.

XX PT Preparation of antisense oligo:nucleotide(s) which lack long runs of consecutive guanosine or inosine - and have specific ratio of residues able to form two or three hydrogen bonds, have greater activity and reduced toxicity, used therapeutically or to modulate growth of cells in culture

XX PS Example 7; Fig 9c; 286pp; English.

XX CC V49008-236 represent antisense oligonucleotides directed against the rb gene. Of these, only oligonucleotides V49008-52 resulted in effective downregulation of negative growth control by rb, while oligonucleotides V49052-236 had little effect. The oligonucleotides exemplify the invention. The specification describes oligonucleotides that contain 8-30 nucleotides, which contain at most 8 nucleotides that can each form three hydrogen bonds to cytosine; do not contain four consecutive nucleotides able to form three H-bonds each to four consecutive cytosines; do not contain two sequences of three consecutive nucleotides each able to form three H-bonds to three consecutive cytosines, and the ratio between residues able to form two H-bonds each (2R) or three such bonds (3R) is given by 2R/3R = 0.33-0.72. The oligonucleotides are used to modulate expression of genes, particularly the genes for p53, ErbB-2, JunB, JunD, TGF-beta 1 or beta 2 to control proliferation of primary cell cultures (e.g. bone marrow stem, liver or kidney cells, osteoclasts, osteoblasts and/or keratinocytes). The oligonucleotides can also be used to analyse function of proteins (by altering their expression or activity) and therapeutically, e.g. in cases of cancer or (targeting TGF) for stimulating the immune system.

XX SQ Sequence 14 BP; 5 A; 1 C; 1 G; 7 T; 0 other;

Query Match 100.0%; Score 7; DB 19; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.8e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttgtgata 7  
|||||||  
DB 1 ttgtgata 7

RESULT 9  
297691/c

ID 297691 standard; DNA; 14 BP.

XX AC 297691;

XX DT 26-APR-2000 (first entry)

XX DE HIV-1 protease gene probe SEQ ID NO:181.

XX KW Human immunodeficiency virus; HIV; protease; probe; detection;  
XX KW drug selected mutation; hybridisation; genotyping; infection;  
XX KW drug resistance; ss.

OS Human immunodeficiency virus type 1.

PN WO9967428-A2.

XX 29-DEC-1999.

XX 22-JUN-1999; 99WO-EP04317.

XX 24-JUN-1998; 98EP-0870143.

XX (INNO-) INNOGENETICS NV.

XX Stuyver L;

XX WPI; 2000-147219/13.

XX Detection of drug-selected mutations in the HIV protease gene used to  
treat HIV infections

XX Claim 3; Page 37; 76pp; English.

XX The present invention describes the detection of drug-selected mutations  
in the HIV protease gene. The method of detection allows the simultaneous  
characterisation of a range of codons involved in drug resistance using  
sets of probes optimised to function together in a reverse-hybridisation  
assay. 297517 to 297997 represent specifically claimed probes for use in  
the assay, and 297479 to 297501 represent specifically claimed HIV  
protease gene polymorphic nucleotide sequences. 297502 to 297515, and  
298004 to 298007, represent PCR primers for the HIV protease gene, and  
297516 represents an HIV protease probe used in an example from the  
present invention. The method, probes and primers can be used for the  
detection of drug-selected mutations in the HIV protease gene. The  
method allows the simultaneous characterisation of a range of codons  
involved in drug resistance. The method may also be used for HIV  
protease genotyping assays. The probes are able to discriminate between  
wild type and mutated protease sequences. The method allows rapid and  
reliable detection of drug-selected mutation in HIV.

XX Sequence 14 BP; 5 A; 1 C; 2 G; 6 T; 0 other;

Query Match 100.0%; Score 7; DB 21; Length 14;

Best Local Similarity 100.0%; Pred. No. 1.8e+04;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7

Db 11 TTTGATA 5

RESULT 10

Q52243 ID Q52243 standard; RNA; 15 BP.

XX AC Q52243;

XX DT 26-MAY-1994 (first entry)

XX Neuroblastoma specific mRNA ribozyme cleavable nucleotide (2670).

XX Multiple drug resistance; mdr-1; ribozyme; membrane protein; liver;  
KW resistance; chemotherapeutic agent; colchicine; doxorubicin; colon;  
KW actinomycin D; vinblastine; small intestine; kidney; adrenal gland;  
KW adenocarcinoma; bowel; transformed phenotype; promyelocytic leukemia;  
KW human; chronic myelogenous leukemia; CML; follicular lymphoma;  
KW B-cell acute lymphocytic leukemia; breast cancer; colon carcinoma;  
KW neuroblastoma; lung cancer; genetic drift; mutation; hammerhead motif;  
KW hairpin; hepatitis delta virus; group I intron; RNaseP; leukaemia; ss.

XX Homo sapiens.

XX OS WO93232057-A.

XX PN

XX

PD 25-NOV-1993.

XX 13-MAY-1993; 93WO-US04573.

XX 14-MAY-1992; 92US-0882822.

PR 14-MAY-1992; 92US-0882885.

PR 26-AUG-1992; 92US-0936110.

PR 26-AUG-1992; 92US-0936421.

PR 26-AUG-1992; 92US-0936531.

PR 26-AUG-1992; 92US-0936532.

PR 07-DEC-1992; 92US-0987131.

PR 19-JAN-1993; 93US-0006122.

PR 19-JAN-1993; 93US-0008910.

XX (RIBO-) RIBOZYME PHARM INC.

XX Draper KG, Thompson JD;

XX WPI; 1993-386203/48.

XX New enzymatic RNA molecules (ribozymes) - which cleave mRNA

XX associated with tumours or mRNA expressed from gene encoding

XX multiple drug resistance

XX Claim 3; Fig 10; 69pp; English.

XX The sequences given in Q51825-2266 represent areas of mRNAs which are  
associated with development or maintenance of chronic myelogenous  
leukemia (CML), promyelocytic leukemia, Burkitt's lymphoma, or  
acute lymphocytic leukemia, follicular lymphoma, B-cell acute  
lymphocytic leukemia, breast cancer, colon carcinoma, neuroblastoma  
and lung cancer. The full length mRNAs containing these target  
sequences, encode aberrant cellular proteins which are able to control  
cellular proliferation and are directly linked to a leukemic  
phenotype. These target sequences are identified by the ribozyme of  
the invention. The ribozymes is formed in a hammerhead motif, but may  
also be formed in the motif of a hairpin, hepatitis delta virus, group  
I intron or RNaseP-like RNA. These ribozymes may be used to inhibit  
the development or expression of a transformed phenotype in man and  
other animals by modulating expression of the corresponding gene.  
Cleavage of target mRNAs expressed in pre-neoplastic and transformed  
cells elicits inhibition of the transformed state. Multiple drug  
resistance (mdr-1) mRNA specific ribozymes remove the mechanism of  
drug resistance used by transformed cells and thus enhances drug  
therapies for tumours. The ribozymes may also be used to study  
genetic drift and mutations within cells.

XX Sequence 15 BP; 4 A; 1 C; 1 G; 9 U; 0 other;

Query Match 100.0%; Score 7; DB 14; Length 15;

Best Local Similarity 42.9%; Pred. No. 1.8e+04;

Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7

Db 1 uuugaua 7

RESULT 11

T56960

ID T56960 standard; RNA; 15 BP.

XX AC T56960;

XX DT 24-APR-1997 (first entry)

XX RSV 1C hammerhead ribozyme target sequence (nt. position 16).

XX Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;

XX gene expression; downregulation; interleukin-5; ICAM-1;

XX intercellular adhesion molecule; rel A; tumour necrosis factor;

Query Match	100.0%;	Score 7;	DB 16;	Length 15;
Best Local Similarity	42.9%;	Pred. No. 1.8e+04;		
Matches	3;	Conservative	4;	Mismatches 0; Indels 0; Gaps 0;
QY	1	tttgata 7		
DB	7	uuugaua 13		
RESULT 12				
T56962				
ID	T56962	standard; RNA; 15 BP.		
XX				
AC	T56962;			
XX				
DT	24-APR-1997	(first entry)		
XX				
DE	RSV LC hammerhead ribozyme target sequence (nt. position 17).			
XX				
KW	Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;			
KW	gene expression; downregulation; interleukin-5; IL-5; ICAM-1;			
KW	intercellular adhesion molecule; rel A; tumour necrosis factor;			
KW	TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;			
KW	translocation; chronic myelogenous leukaemia; CML; cancer;			
KW	Philadelphia chromosome; inflammation; autoimmune disease;			
KW	atherosclerosis; myocardial infarction; stroke; restenosis;			
KW	transplant rejection; rheumatoid arthritis; psoriasis; HIV;			
KW	myocardial ischemia; Kawasaki disease; septic shock; HIV;			
KW	human immunodeficiency virus; acquired immune deficiency syndrome;			
KW	AIDS; ss.			
XX				
OS	Respiratory Syncytial Virus.			
XX				
PN	W09523225-A2.			
XX				
PD	31-AUG-1995.			
XX				
PF	23-FEB-1995;	95WO-IB00156.		
XX				
PR	30-JAN-1995;	95US-0380734.		
PR	23-FEB-1994;	94US-0201109.		
PR	29-MAR-1994;	94US-0218934.		
PR	04-APR-1994;	94US-0222795.		
PR	07-APR-1994;	94US-0224483.		
PR	15-APR-1994;	94US-0227958.		
PR	15-APR-1994;	94US-0228041.		
PR	18-MAY-1994;	94US-0245736.		
PR	06-JUL-1994;	94US-0271280.		
PR	15-AUG-1994;	94US-0291932.		
PR	16-AUG-1994;	94US-0291433.		
PR	17-AUG-1994;	94US-0292620.		
PR	19-AUG-1994;	94US-0293520.		
PR	02-SEP-1994;	94US-0300000.		
PR	08-SEP-1994;	94US-0303039.		
PR	23-SEP-1994;	94US-0311486.		
PR	23-SEP-1994;	94US-0311749.		
PR	28-SEP-1994;	94US-0314397.		
PR	03-OCT-1994;	94US-0316771.		
PR	07-OCT-1994;	94US-0319492.		
PR	11-OCT-1994;	94US-0321993.		
PR	04-NOV-1994;	94US-0334847.		
PR	10-NOV-1994;	94US-0337608.		
PR	28-NOV-1994;	94US-0345516.		
PR	16-DEC-1994;	94US-0357577.		
PR	23-DEC-1994;	94US-0363233.		
XX				
PA	(RIBO-) RIBOZYME PHARM INC.			
XX				
PI	Stinchcomb DT, Chowrira B, Drenzo A, Draper KG, Dudycz LW;			
PI	Grimm S, Karpelsky A, Kislich K, Matulic-Adamic J;			
PI	McSwigler JA, Modak A, Pavco P, Beigelman L, Sullivan SM;			
PI	Swesigler D, Thompson JD, Tracz D, Usman N, Wincott FE;			

```

PI Woolf T;
XX
DR WPI; 1995-351090/45.
XX
PT Ribozymes having modified bases and methods for producing them
PT for use in inhibiting disease related genes
XX
XX
PS Claim 2; Page 269; 407pp; English.
XX
CC The present sequence represents a preferred target sequence for an
CC enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding
CC for a protein of respiratory syncytial virus (RSV) at the
CC nucleotide base position indicated in the DE line. Regions of
CC the mRNA that do not form secondary folding structures and that
CC contain potential hammerhead and hairpin ribozyme cleavage sites
CC were identified by computer analysis. Ribozymes directed against
CC these mRNA sequences were designed and synthesised with modifications
CC that improve their nuclease resistance. The ribozymes cleave the
CC target sequences and can be used for treatment and diagnosis of
CC RSV infection.
XX
XX Sequence 15 BP; 7 A; 0 C; 3 G; 5 U; 0 other;
SQ

```

Query Match 100.0%; Score 7; DB 16; Length 15;  
 Best Local Similarity 42.9%; Pred. No. 1.8e+04;  
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ttgtgata 7
DB 6 uuugaua 12

```

RESULT 13  
 ID T56964 standard; RNA; 15 BP.  
 AC T56964;  
 XX  
 XX  
 DT 24-APR-1997 (first entry)  
 XX  
 DE RSV 1C hammerhead ribozyme target sequence (nt. position 21).  
 XX  
 KW Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;  
 KW gene expression; downregulation; interleukin-5; IL-5; ICAM-1;  
 KW intercellular adhesion molecule; rel A; tumour necrosis factor;  
 KW TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;  
 KW translocation; chronic myelogenous leukaemia; CML; cancer;  
 KW Philadelphia chromosome; inflammation; autoimmune disease;  
 KW atherosclerosis; myocardial infarction; stroke; restenosis;  
 KW transplant rejection; rheumatoid arthritis; psoriasis;  
 KW myocardial ischaemia; Kawasaki disease; septic shock; HIV;  
 KW human immunodeficiency virus; acquired immune deficiency syndrome;  
 KW AIDS; ss.  
 XX  
 OS Respiratory Syncytial Virus.  
 XX  
 PN W09523225-A2.  
 XX  
 PD 31-AUG-1995.  
 XX  
 PF 23-FEB-1995; 95WO-IB00156.  
 XX  
 XX 30-JAN-1995; 95US-0380734.  
 PR 23-FEB-1994; 94US-0201109.  
 PR 29-MAR-1994; 94US-0218934.  
 PR 04-APR-1994; 94US-0222795.  
 PR 07-APR-1994; 94US-0224483.  
 PR 15-APR-1994; 94US-0227958.  
 PR 15-APR-1994; 94US-0228041.  
 PR 18-MAY-1994; 94US-0245736.  
 PR 06-JUL-1994; 94US-0271280.  
 PR 15-AUG-1994; 94US-0291932.

```

PR 16-AUG-1994; 94US-0291433.
PR 17-AUG-1994; 94US-0292620.
PR 19-AUG-1994; 94US-0293520.
PR 02-SEP-1994; 94US-0300000.
PR 08-SEP-1994; 94US-0303039.
PR 23-SEP-1994; 94US-0311486.
PR 28-SEP-1994; 94US-0311749.
PR 03-OCT-1994; 94US-0314397.
PR 07-OCT-1994; 94US-0316771.
PR 11-OCT-1994; 94US-0319492.
PR 04-NOV-1994; 94US-0321993.
PR 10-NOV-1994; 94US-0334847.
PR 28-NOV-1994; 94US-0337608.
PR 16-DEC-1994; 94US-0345516.
PR 23-DEC-1994; 94US-0357577.
XX 94US-0363233.
XX (RIBO-) RIBOZYME PHARM INC.
PA
XX Stinchcomb DT, Chowrira B, Drenzo A, Draper KG, Dudycz LW;
XX Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;
XX McSwiggen JA, Modak A, Pavco P, Beigelman L, Sullivan SM;
XX Svedler D, Thompson JD, Tracz D, Usman N, Wincott FE;
XX Woolf T;
XX
XX WPI; 1995-351090/45.
XX
XX Ribozymes having modified bases and methods for producing them
XX for use in inhibiting disease related genes
XX
XX Claim 2; Page 269; 407pp; English.
XX
XX The present sequence represents a preferred target sequence for an
XX enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding
XX for a protein of respiratory syncytial virus (RSV) at the
XX nucleotide base position indicated in the DE line. Regions of
XX the mRNA that do not form secondary folding structures and that
XX contain potential hammerhead and hairpin ribozyme cleavage sites
XX were identified by computer analysis. Ribozymes directed against
XX these mRNA sequences were designed and synthesised with modifications
XX that improve their nuclease resistance. The ribozymes cleave the
XX target sequences and can be used for treatment and diagnosis of
XX RSV infection.
XX
XX Sequence 15 BP; 6 A; 2 C; 2 G; 5 U; 0 other;
SQ

```

Query Match 100.0%; Score 7; DB 16; Length 15;  
 Best Local Similarity 42.9%; Pred. No. 1.8e+04;  
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ttgtgata 7
DB 2 uuugaua 8

```

RESULT 14  
 T57424/c  
 ID T57424 standard; RNA; 15 BP.  
 XX  
 AC T57424;  
 XX  
 XX 19-MAR-1997 (first entry)  
 DT  
 XX RSV N hammerhead ribozyme target sequence (nt. position 1146).  
 DE  
 XX Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;  
 KW gene expression; downregulation; interleukin-5; IL-5; ICAM-1;  
 KW intercellular adhesion molecule; rel A; tumour necrosis factor;  
 KW TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;  
 KW translocation; chronic myelogenous leukaemia; CML; cancer;  
 KW Philadelphia chromosome; inflammation; autoimmune disease;  
 KW atherosclerosis; myocardial infarction; stroke; restenosis;

KW transplant rejection; rheumatoid arthritis; psoriasis;  
 KW myocardial ischaemia; Kawasaki disease; septic shock; HIV;  
 KW human immunodeficiency virus; acquired immune deficiency syndrome;  
 KW AIDS; ss.

XX Respiratory Syncytial Virus.

OS  
 XX WO9523225-A2.

PN  
 XX 31-AUG-1995.

PD  
 XX 23-FEB-1995; 95WO-IB00156.

PE  
 XX 30-JAN-1995; 95US-0380734.

PF  
 XX 23-FEB-1994; 94US-0201109.

PG  
 XX 29-MAR-1994; 94US-0218934.

PH  
 XX 04-APR-1994; 94US-0222795.

PI  
 XX 07-APR-1994; 94US-0224483.

PJ  
 XX 15-APR-1994; 94US-0227958.

PK  
 XX 15-APR-1994; 94US-0228041.

PL  
 XX 18-MAY-1994; 94US-0245736.

PM  
 XX 06-JUL-1994; 94US-0271280.

PN  
 XX 15-AUG-1994; 94US-0291932.

PO  
 XX 16-AUG-1994; 94US-0291433.

PP  
 XX 17-AUG-1994; 94US-0292620.

PQ  
 XX 19-AUG-1994; 94US-0293520.

PR  
 XX 02-SEP-1994; 94US-0300000.

PS  
 XX 08-SEP-1994; 94US-0303039.

PT  
 XX 23-SEP-1994; 94US-0311486.

PU  
 XX 23-SEP-1994; 94US-0311749.

PV  
 XX 23-SEP-1994; 94US-0314397.

PW  
 XX 03-SEP-1994; 94US-0316771.

PX  
 XX 07-OCT-1994; 94US-0319492.

PY  
 XX 11-OCT-1994; 94US-0321993.

PZ  
 XX 04-NOV-1994; 94US-0334847.

QA  
 XX 10-NOV-1994; 94US-0337608.

QB  
 XX 28-NOV-1994; 94US-0345516.

QC  
 XX 16-DEC-1994; 94US-0357577.

QD  
 XX 23-DEC-1994; 94US-0363233.

QE  
 XX (RIBO-) RIBOZYME PHARM INC.

QF  
 XX Stinchcomb DT, Chowira B, Dorenzo A, Draper KG, Dudycz LW;

QG  
 XX Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;

QH  
 XX McSwiggan JA, Modak A, Pavco P, Belgelman L, Sullivan SM;

QI  
 XX Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;

QJ  
 XX Woolf T;

QK  
 XX WPI; 1995-351090/45.

QL  
 XX

transplant rejection; rheumatoid arthritis; psoriasis;  
 myocardial ischaemia; Kawasaki disease; septic shock; HIV;  
 human immunodeficiency virus; acquired immune deficiency syndrome;  
 AIDS; ss.

Respiratory Syncytial Virus.

WO9523225-A2.

31-AUG-1995.

23-FEB-1995; 95WO-IB00156.

30-JAN-1995; 95US-0380734.

23-FEB-1994; 94US-0201109.

29-MAR-1994; 94US-0218934.

04-APR-1994; 94US-0222795.

07-APR-1994; 94US-0224483.

15-APR-1994; 94US-0227958.

15-APR-1994; 94US-0228041.

18-MAY-1994; 94US-0245736.

06-JUL-1994; 94US-0271280.

15-AUG-1994; 94US-0291932.

16-AUG-1994; 94US-0291433.

17-AUG-1994; 94US-0292620.

19-AUG-1994; 94US-0293520.

02-SEP-1994; 94US-0300000.

08-SEP-1994; 94US-0303039.

23-SEP-1994; 94US-0311486.

23-SEP-1994; 94US-0311749.

03-SEP-1994; 94US-0314397.

07-OCT-1994; 94US-0316771.

11-OCT-1994; 94US-0321993.

04-NOV-1994; 94US-0334847.

10-NOV-1994; 94US-0337608.

28-NOV-1994; 94US-0345516.

16-DEC-1994; 94US-0357577.

23-DEC-1994; 94US-0363233.

(RIBO-) RIBOZYME PHARM INC.

Stinchcomb DT, Chowira B, Dorenzo A, Draper KG, Dudycz LW;  
 Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;  
 McSwiggan JA, Modak A, Pavco P, Belgelman L, Sullivan SM;  
 Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;  
 Woolf T;

WPI; 1995-351090/45.

Ribozymes having modified bases and methods for producing them -  
 for use in inhibiting disease related genes.

Claim 2; Page 276; 407pp; English.

The present sequence represents a preferred target sequence for an  
 enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding  
 for a protein of respiratory syncytial virus (RSV) at the  
 nucleotide base position indicated in the DE line. Regions of  
 the mRNA that do not form secondary folding structures and that  
 contain potential hammerhead and hairpin ribozyme cleavage sites  
 were identified by computer analysis. Ribozymes directed against  
 these mRNA sequences were designed and synthesised with modifications  
 that improve their nuclease resistance. The ribozymes cleave the  
 target sequences and can be used for treatment and diagnosis of  
 RSV infection.

Sequence 15 BP; 7 A; 3 C; 2 U; 0 other;

Query Match 100.0%; Score 7; DB 16; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttgtgata 7  
 Db 13 TTTGATA 7

RESULT 15  
 T57425/c  
 ID T57425 standard; RNA; 15 BP.  
 XX  
 AC T57425;  
 XX  
 DT 19-MAR-1997 (first entry)  
 XX  
 DE RSV N hammerhead ribozyme target sequence (nt. position 1148).  
 XX  
 KW Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;  
 KW gene expression; downregulation; interleukin-5; IL-5; ICAM-1;  
 KW intercellular adhesion molecule; rel A; tumour necrosis factor;  
 KW TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;  
 KW translocation; chronic myelogenous leukaemia; CML; cancer;  
 KW Philadelphia chromosome; inflammation; autoimmune disease;  
 KW atherosclerosis; myocardial infarction; stroke; restenosis;  
 KW transplant rejection; rheumatoid arthritis; psoriasis;  
 KW myocardial ischaemia; Kawasaki disease; septic shock; HIV;  
 KW human immunodeficiency virus; acquired immune deficiency syndrome;  
 KW AIDS; ss.

XX  
 OS Respiratory Syncytial Virus.  
 XX  
 PN WO9523225-A2.  
 XX  
 PD 31-AUG-1995.  
 XX  
 PF 23-FEB-1995; 95WO-IB00156.  
 XX  
 PR 30-JAN-1995; 95US-0380734.  
 PR 23-FEB-1994; 94US-0201109.  
 PR 29-MAR-1994; 94US-0218934.  
 PR 04-APR-1994; 94US-0222795.  
 PR 07-APR-1994; 94US-0224483.  
 PR 15-APR-1994; 94US-0227958.  
 PR 15-APR-1994; 94US-0228041.  
 PR 18-MAY-1994; 94US-0245736.  
 PR 06-JUL-1994; 94US-0271280.  
 PR 15-AUG-1994; 94US-0291932.  
 PR 16-AUG-1994; 94US-0291433.  
 PR 17-AUG-1994; 94US-0292620.  
 PR 19-AUG-1994; 94US-0293520.  
 PR 02-SEP-1994; 94US-0300000.  
 PR 08-SEP-1994; 94US-0303039.  
 PR 23-SEP-1994; 94US-0311486.  
 PR 23-SEP-1994; 94US-0311749.  
 PR 28-SEP-1994; 94US-0314397.  
 PR 03-OCT-1994; 94US-0316771.  
 PR 07-OCT-1994; 94US-0319492.  
 PR 11-OCT-1994; 94US-0321993.  
 PR 04-NOV-1994; 94US-0334847.  
 PR 10-NOV-1994; 94US-0337608.  
 PR 28-NOV-1994; 94US-0345516.  
 PR 16-DEC-1994; 94US-0357577.  
 PR 23-DEC-1994; 94US-0363233.

XX  
 (RIBO-) RIBOZYME PHARM INC.  
 XX  
 PA Stinchcomb DT, Chowira B, Dorenzo A, Draper KG, Dudycz LW;  
 PI Grimm S, Karpeisky A, Kisich K, Matulic-Adamic J;  
 PI McSwiggan JA, Modak A, Pavco P, Belgelman L, Sullivan SM;  
 PI Sweedler D, Thompson JD, Tracz D, Usman N, Wincott FE;  
 PI Woolf T;  
 XX  
 DR WPI; 1995-351090/45.  
 XX

PT Ribozymes having modified bases and methods for producing them  
PT , for use in inhibiting disease related genes  
XX  
XX Claim 2; Page 276; 407pp; English.  
XX  
CC The present sequence represents a preferred target sequence for an  
CC enzymatic nucleic acid (i.e. a ribozyme) which cleaves mRNA coding  
CC for a protein of respiratory syncytial virus (RSV) at the  
CC nucleotide base position indicated in the DE line. Regions of  
CC the mRNA that do not form secondary folding structures and that  
CC contain potential hammerhead and hairpin ribozyme cleavage sites  
CC were identified by computer analysis. Ribozymes directed against  
CC these mRNA sequences were designed and synthesised with modifications  
CC that improve their nuclease resistance. The ribozymes cleave the  
CC target sequences and can be used for treatment and diagnosis of  
CC RSV infection.  
XX  
SQ Sequence 15 BP; 6 A; 4 C; 2 G; 3 U; 0 other;

Query Match 100.0%; Score 7; DB 16; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.8e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgtata 7  
| | | | |  
Db 11 TTTGATA 5

Search completed: March 27, 2001, 08:22:59  
Job time: 5301 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:22 ; Search time 132.3 Seconds  
(without alignments)  
8.527 Million cell updates/sec

Title: US-09-380-826A-5

Perfect score: 7

Sequence: 1 ttgata 7

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	7	100.0	11	1	US-08-086-634-6
C 2	7	100.0	11	1	US-08-086-634-7
C 3	7	100.0	13	1	US-08-166-664-5
C 4	7	100.0	13	1	US-08-271-880A-217
C 5	7	100.0	13	2	US-08-910-408-217
C 6	7	100.0	13	3	US-09-249-213-217
C 7	7	100.0	14	1	US-08-271-880A-212
C 8	7	100.0	14	2	US-08-910-408-212
C 9	7	100.0	14	3	US-09-249-215-212
C 10	7	100.0	15	1	US-08-373-124A-90
C 11	7	100.0	15	1	US-08-291-932A-129
C 12	7	100.0	15	1	US-08-291-932A-130
C 13	7	100.0	15	1	US-08-291-932A-131
C 14	7	100.0	15	1	US-08-334-847-8
C 15	7	100.0	15	1	US-08-334-847-9
C 16	7	100.0	15	1	US-08-334-847-10
C 17	7	100.0	15	1	US-08-334-847-11
C 18	7	100.0	15	1	US-08-334-847-12
C 19	7	100.0	15	1	US-08-334-847-13
C 20	7	100.0	15	1	US-08-435-628-90
C 21	7	100.0	15	2	US-08-585-684B-113
C 22	7	100.0	15	2	US-08-585-684B-114
C 23	7	100.0	15	2	US-08-585-684B-115
C 24	7	100.0	15	2	US-08-585-684B-116
C 25	7	100.0	15	2	US-08-585-684B-117
C 26	7	100.0	15	2	US-08-585-684B-118
C 27	7	100.0	15	2	US-08-585-684B-119
C 28	7	100.0	15	2	US-08-585-684B-120

C 29	7	100.0	15	2	US-08-585-684B-2320
C 30	7	100.0	15	2	US-08-585-684B-2321
C 31	7	100.0	15	2	US-08-634-350-25
C 32	7	100.0	16	3	US-08-930-917A-7
C 33	7	100.0	16	3	US-08-930-917A-8
C 34	7	100.0	17	1	US-08-373-124A-520
C 35	7	100.0	17	1	US-08-373-124A-522
C 36	7	100.0	17	1	US-08-373-124A-521
C 37	7	100.0	17	1	US-08-373-124A-1563
C 38	7	100.0	17	1	US-08-373-124A-2411
C 39	7	100.0	17	1	US-08-373-124A-2413
C 40	7	100.0	17	1	US-08-373-124A-2565
C 41	7	100.0	17	1	US-08-373-124A-2567
C 42	7	100.0	17	1	US-08-435-628-520
C 43	7	100.0	17	1	US-08-435-628-522
C 44	7	100.0	17	1	US-08-435-628-1561
C 45	7	100.0	17	1	US-08-435-628-1563

#### ALIGNMENTS

RESULT 1

US-08-634-6/c

; Sequence 6, Application US/08086634

; Patent No. 5516634

; GENERAL INFORMATION:

; APPLICANT: NEWMAN, Peter J.

; APPLICANT: SANTOSO, Sentot Saleh

; TITLE OF INVENTION: Molecular Basis of the

; TITLE OF INVENTION: Human Platelet Bra/Brb Alloantigen System and

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: P. O. Box 1497

; CITY: Madison

; STATE: Wisconsin

; COUNTRY: USA

; ZIP: 53701-1497

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WordPerfect, Version 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/086,634

; FILING DATE: 30-June-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Scanlon, William J.

; REGISTRATION NUMBER: 30,136

; REFERENCE/DOCKET NUMBER: 30383/132

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (608) 258-4284

; TELEFAX: (608) 258-4258

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

US-08-086-634-6

Query Match 100.0%; Score 7; DB 1; Length 11;

Best Local Similarity 100.0%; Pred No. 7.1e+03;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgata 7

Db 7 TTTGATA 1

RESULT 2  
US-08-086-634-7/c  
; Sequence 7, Application US/08086634  
; Patent No. 5516634  
; GENERAL INFORMATION:  
; APPLICANT: NEWMAN, Peter J.  
; APPLICANT: SANTOSO, Sentot Saleh  
; TITLE OF INVENTION: Molecular Basis of the  
; TITLE OF INVENTION: Human Platelet Bra/Brb Alloantigen System and  
; TITLE OF INVENTION: Applications Thereof  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: P. O. Box 1497  
; CITY: Madison  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53701-1497  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect, Version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/086.634  
; FILING DATE: 30-June-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Scanlon, William J.  
; REGISTRATION NUMBER: 30,136  
; REFERENCE/DOCKET NUMBER: 30383/132  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (608) 258-4284  
; TELEFAX: (608) 258-4258  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
US-08-086-634-7

Query Match 100.0%; Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 7.1e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
      |||||  
DB 7 TTTGATA 1

RESULT 3  
US-08-166-664-5/c  
; Sequence 5, Application US/08166664  
; Patent No. 5646020  
; GENERAL INFORMATION:  
; APPLICANT: James A. McSwiggen  
; APPLICANT: J. Anthony Mamone  
; TITLE OF INVENTION: HAMMERHEAD RIBOZYMES FOR  
; TITLE OF INVENTION: PREFERRED TARGETS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 611 West Sixth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/166,664  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/884,074  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 197/062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-166-664-5

Query Match 100.0%; Score 7; DB 1; Length 13;  
Best Local Similarity 100.0%; Pred. No. 7e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
      |||||  
DB 10 TTTGATA 4

RESULT 4  
US-08-271-880A-217/c  
; Sequence 217, Application US/08271880A  
; Patent No. 5693535  
; GENERAL INFORMATION:  
; APPLICANT: Kenneth G. Draper  
; APPLICANT: Bharat Chowhira  
; APPLICANT: James McSwiggen  
; APPLICANT: Dan T. Stinchcomb  
; APPLICANT: James D. Thompson  
; TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING  
; TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS  
; TITLE OF INVENTION: REPLICATION  
; NUMBER OF SEQUENCES: 232  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/271,880A  
; FILING DATE: July 7, 1994  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: described below:  
; APPLICATION NUMBER: 08/103,243  
; FILING DATE: August 6, 1993  
; APPLICATION NUMBER: 07/882,886  
; FILING DATE: May 14, 1992

two

ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 206/116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 217:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-271-880A-217

Query Match 100.0%; Score 7; DB 1; Length 13;  
Best Local Similarity 100.0%; Pred. No. 7e+03; 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
DB 10 TTTGATA 4

RESULT 5  
US-08-910-408-217/c  
Sequence 217, Application US/08910408  
Patent No. 5972704  
GENERAL INFORMATION:  
APPLICANT: Kenneth G. Draper  
APPLICANT: Bharat Chowrira  
APPLICANT: James McSwiggen  
APPLICANT: Dan T. Stinchcomb  
APPLICANT: James D. Thompson  
TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING  
TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS  
NUMBER OF SEQUENCES: 232  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,408  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/271,880  
FILING DATE: July 7, 1994  
APPLICATION NUMBER: 08/103,243  
FILING DATE: August 6, 1993  
APPLICATION NUMBER: 07/882,886  
FILING DATE: May 14, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 206/116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 217:

SEQUENCE CHARACTERISTICS:  
LENGTH: 13 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-910-408-217

Query Match 100.0%; Score 7; DB 2; Length 13;  
Best Local Similarity 100.0%; Pred. No. 7e+03; 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
DB 10 TTTGATA 4

RESULT 6  
US-09-249-215-217/c  
Sequence 217, Application US/09249215  
Patent No. 6159692  
GENERAL INFORMATION:  
APPLICANT: Kenneth G. Draper  
APPLICANT: Bharat Chowrira  
APPLICANT: James McSwiggen  
APPLICANT: Dan T. Stinchcomb  
APPLICANT: James D. Thompson  
TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING  
TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS  
NUMBER OF SEQUENCES: 232  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/249,215  
FILING DATE: 12-Feb-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/910,408  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/103,243  
FILING DATE: August 6, 1993  
APPLICATION NUMBER: 07/882,886  
FILING DATE: May 14, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 206/116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 217:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 217:  
US-09-249-215-217

us-09-380-826a-5.rni

Wed Mar 28 14:03:22 2001

Query Match 100.0%; Score 7; DB 3; Length 13;  
Best Local Similarity 100.0%; Pred. No. 7e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7  
Db 10 TTTGATA 4

RESULT 7

US-08-271-880A-212  
; Sequence 212, Application US/08271880A

; Patent No. 5693535

; GENERAL INFORMATION:

; APPLICANT: Kenneth G. Draper

; APPLICANT: Bharat Chowhira

; APPLICANT: James McSwiggen

; APPLICANT: Dan T. Stinchcomb

; APPLICANT: James D. Thompson

; TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING

; TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS

; NUMBER OF SEQUENCES: 232

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; CITY: Suite 4700

; STATE: Los Angeles

; COUNTRY: California

; ZIP: U.S.A.

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: Storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/271,880A

; FILING DATE: July 7, 1994

; PRIOR APPLICATION DATA:

; APPLICATION DATA: including application

; PRIOR APPLICATION DATA: described below:

; APPLICATION NUMBER: 08/103,243

; FILING DATE: August 6, 1993

; APPLICATION NUMBER: 07/882,886

; FILING DATE: May 14, 1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Warburg, Richard

; REGISTRATION NUMBER: 32,327

; REFERENCE/DOCKET NUMBER: 206/116

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; INFORMATION FOR SEQ ID NO: 212:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 14 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-271-880A-212

Query Match 100.0%; Score 7; DB 1; Length 14;  
Best Local Similarity 100.0%; Pred. No. 7e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7  
Db 3 TTTGATA 9

RESULT 9

US-09-249-215-212

; Sequence 212, Application US/09249215

; Patent No. 6159692

; GENERAL INFORMATION:

; APPLICANT: Kenneth G. Draper

; APPLICANT: Bharat Chowhira

; APPLICANT: James McSwiggen

; APPLICANT: Dan T. Stinchcomb

RESULT 8

US-08-910-408-212

; Sequence 212, Application US/08910408

; Patent No. 5972704

; GENERAL INFORMATION:

; APPLICANT: Kenneth G. Draper

; APPLICANT: Bharat Chowhira

; APPLICANT: James McSwiggen

; APPLICANT: Dan T. Stinchcomb

; APPLICANT: James D. Thompson

; TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING

; TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS

; NUMBER OF SEQUENCES: 232

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; CITY: Suite 4700

; STATE: Los Angeles

; COUNTRY: California

; ZIP: U.S.A.

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: Storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/910,408

; FILING DATE: May 14, 1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/271,880

; FILING DATE: July 7, 1994

; APPLICATION NUMBER: 08/103,243

; FILING DATE: August 6, 1993

; APPLICATION NUMBER: 07/882,886

; FILING DATE: May 14, 1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Warburg, Richard

; REGISTRATION NUMBER: 32,327

; REFERENCE/DOCKET NUMBER: 206/116

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; INFORMATION FOR SEQ ID NO: 212:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 14 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-910-408-212

Query Match 100.0%; Score 7; DB 2; Length 14;  
Best Local Similarity 100.0%; Pred. No. 7e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7  
Db 3 TTTGATA 9

```

; James D. Thompson
; TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
; HUMAN IMMUNODEFICIENCY VIRUS
; REPLICATION
;
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ Version 1.5
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,215
; FILING DATE: 12-Feb-1999
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/910,408
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/103,243
; FILING DATE: August 6, 1993
; APPLICATION NUMBER: 07/882,886
; FILING DATE: May 14, 1992
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 206/116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
;
; INFORMATION FOR SEQ ID NO: 212:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 212:
;
; US-09-249-215-212
;
; Query Match 100.0%; Score 7; DB 3; Length 14;
; Best Local Similarity 100.0%; Pred. No. 7e+03;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 ttgtgata 7
; Db 3 TTTGATA 9
;
; RESULT 10
; US-08-373-124A-90/c
; Sequence 90, Application US/08373124A
; Patent No. 5646042
;
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; Suite 4700

```

```

; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,124A
; FILING DATE: January 13, 1995
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
;
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-373-124A-90
;
; Query Match 100.0%; Score 7; DB 1; Length 15;
; Best Local Similarity 100.0%; Pred. No. 7e+03;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 ttgtgata 7
; Db 11 TTTGATA 5
;
; RESULT 11
; US-08-291-932A-129
; Sequence 129, Application US/08291932A
; Patent No. 5658780
;
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth G.
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: RIBOZYME TREATMENT OF
; DISEASES OR CONDITIONS
; TITLE OF INVENTION: RELATED TO LEVELS OF
; TITLE OF INVENTION: NF-KB
; NUMBER OF SEQUENCES: 830
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;
; COMPUTER: IBM Compatible

```

OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291.932A  
FILING DATE: August 15, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 08/245.466  
FILING DATE: May 18, 1994  
APPLICATION NUMBER: 07/987.132  
FILING DATE: December 7, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 208/157  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 130:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-291-932A-130

Query Match 100.0%; Score 7; DB 1; Length 15;  
Best Local Similarity 42.9%; Pred. No. 7e+03;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7  
Db 7 UUUGAUA 13

RESULT 13  
US-08-291-932A-131  
Sequence 131, Application US/08291932A  
Patent No. 5658780  
GENERAL INFORMATION:  
APPLICANT: Stinchcomb, Dan T.  
APPLICANT: Draper, Kenneth G.  
APPLICANT: McSwiggen, James  
TITLE OF INVENTION: RIBOZYME TREATMENT OF  
TITLE OF INVENTION: DISEASES OR CONDITIONS  
TITLE OF INVENTION: RELATED TO LEVELS OF  
TITLE OF INVENTION: NF-KB  
NUMBER OF SEQUENCES: 830  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291.932A  
FILING DATE: August 15, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 08/245.466  
FILING DATE: May 18, 1994  
APPLICATION NUMBER: 07/987.132  
FILING DATE: December 7, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 208/157  
TELECOMMUNICATION INFORMATION:

OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291.932A  
FILING DATE: August 15, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 08/245.466  
FILING DATE: May 18, 1994  
APPLICATION NUMBER: 07/987.132  
FILING DATE: December 7, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 208/157  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 129:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-291-932A-129

Query Match 100.0%; Score 7; DB 1; Length 15;  
Best Local Similarity 42.9%; Pred. NO. 7e+03;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtata 7  
Db 9 UUUGAUA 15

RESULT 12  
US-08-291-932A-130  
Sequence 130, Application US/08291932A  
Patent No. 5658780  
GENERAL INFORMATION:  
APPLICANT: Stinchcomb, Dan T.  
APPLICANT: Draper, Kenneth G.  
APPLICANT: McSwiggen, James  
TITLE OF INVENTION: RIBOZYME TREATMENT OF  
TITLE OF INVENTION: DISEASES OR CONDITIONS  
TITLE OF INVENTION: RELATED TO LEVELS OF  
TITLE OF INVENTION: NF-KB  
NUMBER OF SEQUENCES: 830  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291.932A  
FILING DATE: August 15, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:

Two

Two

TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 131:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-291-932A-131

Query Match 100.0%; Score 7; DB 1; Length 15;  
Best Local Similarity 42.9%; Pred. No. 7e+03;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
: : : : :  
Db 6 UUUGAUA 12

RESULT 14  
US-08-334-847-8  
Sequence 8, Application US/08334847  
Patent No. 5693532  
GENERAL INFORMATION:  
APPLICANT: McSwiggen, James  
APPLICANT: Draper, Kenneth  
APPLICANT: Pavco, Pam  
APPLICANT: Woolf, Tod  
TITLE OF INVENTION: METHOD AND REAGENT FOR  
INHIBITING RESPIRATORY  
TITLE OF INVENTION: INHIBITING RESPIRATORY  
NUMBER OF SEQUENCES: 909  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/334,847  
FILING DATE: No. 5693532ember 4, 1994  
PRIOR APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 209/032  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-334-847-8

Query Match 100.0%; Score 7; DB 1; Length 15;  
Best Local Similarity 42.9%; Pred. No. 7e+03;

Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ttgtgata 7  
: : : : :  
Db 7 UUUGAUA 13

RESULT 15  
US-08-334-847-9  
Sequence 9, Application US/08334847  
Patent No. 5693532  
GENERAL INFORMATION:  
APPLICANT: McSwiggen, James  
APPLICANT: Draper, Kenneth  
APPLICANT: Pavco, Pam  
APPLICANT: Woolf, Tod  
TITLE OF INVENTION: METHOD AND REAGENT FOR  
INHIBITING RESPIRATORY  
TITLE OF INVENTION: INHIBITING RESPIRATORY  
NUMBER OF SEQUENCES: 909  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/334,847  
FILING DATE: No. 5693532ember 4, 1994  
PRIOR APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 209/032  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-334-847-9

Query Match 100.0%; Score 7; DB 1; Length 15;  
Best Local Similarity 42.9%; Pred. No. 7e+03;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
: : : : :  
Db 6 UUUGAUA 12

Search completed: March 27, 2001, 08:19:22  
Job time: 5149 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:21 ; Search time 2517.78 Seconds  
(without alignments)  
19.482 Million cell updates/sec

Title: US-09-380-826A-5  
Perfect score: 7  
Sequence: 1 ttgata 7

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_estl1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estrol1:\*  
75: em\_estrol2:\*  
76: em\_estrol3:\*  
77: em\_estrol4:\*  
78: em\_estrol5:\*  
79: em\_estrol6:\*  
80: em\_estrol7:\*  
81: em\_estrol8:\*  
82: em\_estrol9:\*  
83: em\_estrol10:\*  
84: em\_estrol11:\*  
85: em\_estrol12:\*  
86: em\_estrol13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
118: em\_estp17:\*  
119: em\_estp18:\*  
120: em\_estp19:\*  
121: em\_estp20:\*  
122: em\_estp21:\*  
123: em\_estp22:\*  
124: em\_estp23:\*  
125: em\_estp24:\*  
126: em\_estp25:\*  
127: em\_estp26:\*  
128: em\_estp27:\*  
129: em\_estp28:\*  
130: em\_estp29:\*  
131: em\_estp30:\*  
132: em\_estp31:\*  
133: em\_estp32:\*  
134: em\_estp33:\*  
135: em\_estp34:\*  
136: em\_estp35:\*  
137: em\_estp36:\*  
138: em\_estp37:\*  
139: em\_estp38:\*  
140: em\_estp39:\*  
141: em\_estp40:\*  
142: em\_estp41:\*  
143: em\_estp42:\*  
144: em\_estp43:\*  
145: em\_estp44:\*  
146: em\_estp45:\*  
147: em\_estp46:\*  
148: em\_estp47:\*  
149: em\_estp48:\*  
150: em\_estp49:\*  
151: em\_estp50:\*  
152: em\_estp51:\*  
153: em\_estp52:\*  
154: em\_estp53:\*  
155: em\_estp54:\*  
156: em\_estp55:\*  
157: em\_estp56:\*  
158: em\_estp57:\*  
159: em\_estp58:\*  
160: em\_estp59:\*  
161: em\_estp60:\*  
162: em\_estp61:\*  
163: em\_estp62:\*  
164: em\_estp63:\*  
165: em\_estp64:\*  
166: em\_estp65:\*  
167: em\_estp66:\*  
168: em\_estp67:\*  
169: em\_estp68:\*  
170: em\_estp69:\*  
171: em\_estp70:\*  
172: em\_estp71:\*  
173: em\_estp72:\*  
174: em\_estp73:\*  
175: em\_estp74:\*  
176: em\_estp75:\*  
177: em\_estp76:\*  
178: em\_estp77:\*  
179: em\_estp78:\*  
180: em\_estp79:\*  
181: em\_estp80:\*  
182: em\_estp81:\*  
183: em\_estp82:\*  
184: em\_estp83:\*  
185: em\_estp84:\*  
186: em\_estp85:\*  
187: em\_estp86:\*  
188: em\_estp87:\*  
189: em\_estp88:\*

190: gb\_gss25:\*  
191: gb\_gss26:\*  
192: gb\_gss27:\*  
193: gb\_gss28:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	7	100.0	20	170	AZ323820	AZ323820 IM0045A01
C 2	7	100.0	22	170	AZ323820	AZ323820 IM0035C07
C 3	7	100.0	22	173	AZ502952	AZ502952 IM0342C04
C 4	7	100.0	23	171	AZ345476	AZ345476 IM0080P11
C 5	7	100.0	24	171	AZ379215	AZ379215 IM0134M14
C 6	7	100.0	24	173	AZ480650	AZ480650 IM0302P17
C 7	7	100.0	25	28	AU008929	AU008929 AU008929
C 8	7	100.0	25	172	AZ404619	AZ404619 IM0173L20
9	7	100.0	26	89	AW333097	AW333097 SI7C3 AGS
10	7	100.0	27	140	D18735	D18735 MUSGS01797
11	7	100.0	27	170	AZ328549	AZ328549 IM0052D24
12	7	100.0	27	170	AZ335603	AZ335603 IM0055E13
C 13	7	100.0	30	28	AU007442	AU007442 AU007442
C 14	7	100.0	30	137	BE911076	BE911076 601662175
C 15	7	100.0	30	139	C20899	C20899 HUMGS000497
C 16	7	100.0	31	16	A1140482	A1140482 ow82d12.s
C 17	7	100.0	31	27	A1973656	A1973656 sd07h11.y
C 18	7	100.0	32	28	AU009849	AU009849 AU009849
19	7	100.0	32	140	D18230	D18230 MUSGS00309
20	7	100.0	32	172	AZ447488	AZ447488 IM0244K07
21	7	100.0	33	145	R98634	R98634 yq69h06.r1
22	7	100.0	33	170	AZ308864	AZ308864 IM0012M13
23	7	100.0	33	172	AZ411435	AZ411435 IM0184C05
C 24	7	100.0	34	1	AA009774	AA009774 z104006.s
C 25	7	100.0	34	148	A0025283	A0025283 EP(3)3105
C 26	7	100.0	34	171	AZ387837	AZ387837 IM0147H19
C 27	7	100.0	34	173	AZ491530	AZ491530 IM0325H09
C 28	7	100.0	35	28	AU008005	AU008005 AU008005
C 29	7	100.0	36	28	AU012250	AU012250 AU012250
C 30	7	100.0	36	28	AU012255	AU012255 AU012255
C 31	7	100.0	37	10	AA692338	AA692338 vt20f07.r
C 32	7	100.0	37	13	AA916682	AA916682 on14610.s
33	7	100.0	37	172	AZ423769	AZ423769 IM0203G22
34	7	100.0	37	172	AZ429862	AZ429862 IM0214I05
35	7	100.0	37	173	AZ481957	AZ481957 IM0306F12
C 36	7	100.0	38	112	HS0002249	HS0002249 HOMO_SBP1
C 37	7	100.0	38	137	BE886395	BE886395 601509796
C 38	7	100.0	38	139	C21092	C21092 HUMGS000260
39	7	100.0	39	141	H38033	H38033 YP58a09.s1
C 40	7	100.0	39	172	AZ402088	AZ402088 IM0169B16
41	7	100.0	39	172	AZ407162	AZ407162 IM0176A13
42	7	100.0	40	3	AA213154	AA213154 mw85c08.r
43	7	100.0	40	4	AA226145	AA226145 nc09f09.r
44	7	100.0	40	12	AA846437	AA846437 a185c09.s
45	7	100.0	40	12	AA846437	AA846437 a185c09.s

ALIGNMENTS

RESULT 1  
AZ323820/c 20 bp DNA  
LOCUS 1M0045A01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0045A01 R, DNA sequence.  
ACCESSION AZ323820.1 GI:10378917  
VERSION AZ323820.1  
KEYWORDS GSS.  
SOURCE house mouse.

```

ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 20) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE         Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL       Unpublished (2000)
COMMENT       Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0045 row: A column: 01
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 20.
FEATURES      Location/Qualifiers
1..20
source        /organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0045A01"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT    10 a      4 c      3 g      3 t
ORIGIN
Query Match    100.0%; Score 7; DB 170; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
|||||
Db 11 TTTGATA 5

RESULT 2
AZ329387/c
LOCUS      AZ329387      22 bp      DNA      GSS      29-SEP-2000
DEFINITION 1M0053C07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0053C07 R, DNA sequence.
ACCESSION  AZ329387
VERSION     AZ329387.1 GI:10390050
KEYWORDS    GSS.

```

```

SOURCE        house mouse.
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 22) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE         Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL       Unpublished (2000)
COMMENT       Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0053 row: C column: 07
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 22.
FEATURES      Location/Qualifiers
1..22
source        /organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0053C07"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT    8 a      4 c      3 g      7 t
ORIGIN
Query Match    100.0%; Score 7; DB 170; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7
|||||
Db 21 TTTGATA 15

RESULT 3
AZ502952/c
LOCUS      AZ502952      22 bp      DNA      GSS      05-OCT-2000
DEFINITION 1M0342C04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0342C04 R, DNA sequence.
ACCESSION  AZ502952
VERSION     AZ502952.1 GI:10684268

```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Matches  
QY  
Db  
RESULT  
LOCUS  
DEFINITION  
ACCESSION

AZ345476.1 GI:10424713  
GSS  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 23)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0080 row: P column: 11  
Seq primer: CGTTGTAACGACGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 23.  
Location/Qualifiers  
1. .23  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0080P11"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gll4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

100.0%; Score 7; DB 171; Length 23;  
100.0%; Pred. No. 2.2e+05;  
7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ttgtgata 7  
|||||  
23 TTTGATA 17  
12 a 1 c 0 g 10 t  
BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 7; DB 171; Length 23;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ttgtgata 7  
Db 23 TTTGATA 17  
RESULT 5  
AZ379215/c  
LOCUS  
DEFINITION  
AZ379215 24 bp DNA  
1M0134M14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0134M14 F, DNA sequence.

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Matches  
QY  
Db  
RESULT  
LOCUS  
DEFINITION  
ACCESSION

AZ345476.1 GI:10424713  
GSS  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 22)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0342 row: C column: 04  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 22.  
Location/Qualifiers  
1. .22  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0342C04"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gll4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

100.0%; Score 7; DB 173; Length 22;  
100.0%; Pred. No. 2.2e+05;  
7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ttgtgata 7  
|||||  
9 TTTGATA 3  
10 a 5 c 1 g .6 t  
BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 7; DB 173; Length 22;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ttgtgata 7  
Db 9 TTTGATA 3  
RESULT 4  
AZ345476/c  
LOCUS  
DEFINITION  
ACCESSION

AZ345476 23 bp DNA  
1M0080P11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0080P11 F, DNA sequence.

ACCESSION AZ379215  
 VERSION AZ379215.1 GI:10492915  
 KEYWORDS GSS.  
 SOURCE house musculus  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL plasmid inserts  
 COMMENT Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0134 row: M column: 14  
 Seq primer: CGTTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 24.

FEATURES Location/Qualifiers

1..24  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0134M14"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (g114732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 8 a 4 c 4 g 8 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 171; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
 |||||  
 Db 12 TTTGATA 6

RESULT 6  
 AZ480650/c 24 bp DNA GSS 04-OCT-2000  
 LOCUS 1M0302P17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 COMMENT

clone UUGC1M0302P17 F, DNA sequence.  
 AZ480650  
 AZ480650.1 GI:10641715  
 GSS.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 24)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0302 row: P column: 17  
 Seq primer: CGTTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 24.

FEATURES Location/Qualifiers

1..24  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0302P17"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (g114732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 11 a 4 c 1 g 8 t  
 ORIGIN

Query Match 100.0%; Score 7; DB 173; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
 |||||  
 Db 8 TTTGATA 2

RESULT 7  
 AU008929/c 25 bp mRNA EST 31-JUL-1998  
 LOCUS AU008929

DEFINITION AU008929 Schizosaccharomyces pombe late log phase cDNA  
Schizosaccharomyces pombe cDNA clone spc04231, mRNA sequence.  
ACCESSION AU008929  
VERSION AU008929.1 GI:3345387  
KEYWORDS EST.  
SOURCE fission yeast.  
ORGANISM Schizosaccharomyces pombe  
Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.  
REFERENCE 1 (bases 1 to 25)  
AUTHORS Morimyo,M. and Mita,K.  
TITLE Identification of expressed sequence tags of Schizosaccharomyces pombe  
JOURNAL Unpublished (1998)  
COMMENT Contact: Mitsuoki Morimyo  
Genome Research Group  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan  
Email: morimyo@nirs.go.jp.  
Location/Qualifiers  
FEATURES  
1. 25  
/organism="Schizosaccharomyces pombe"  
/strain="972"  
/db\_xref="taxon:4896"  
/clone="spc04231"  
/clone\_lib="Schizosaccharomyces pombe late log phase cDNA"  
/sex="h minus"  
/note="Vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"  
BASE COUNT 16 a 3 c 2 g 4 t  
ORIGIN

Query Match 100.0%; Score 7; DB 28; Length 25;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
|||||  
DB 11 TTTGATA 5

RESULT 8  
AZ0404619 25 bp DNA GSS 03-OCT-2000  
LOCUS IM0173L20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0173L20 F, DNA sequence.  
DEFINITION AZ0404619  
ACCESSION AZ0404619.1 GI:10528632  
VERSION  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 25)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D.,Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0173 row: L column: 20  
Seq primer: CGTTGTAAACGACGCGCCACT  
Class: plasmid ends  
High quality sequence stop: 25.  
Location/Qualifiers  
1. 25  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0173L20"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
BASE COUNT 6 a 3 c 4 g 12 t  
ORIGIN

Query Match 100.0%; Score 7; DB 172; Length 25;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7  
|||||  
DB 4 TTTGATA 10

RESULT 9  
AW333097 26 bp mRNA EST 31-JAN-2000  
LOCUS SL7C3 AGS-1 Pneumocystis carinii f. sp. carinii CDNA 3', mRNA sequence.  
DEFINITION AW333097  
ACCESSION AW333097.1 GI:6829454  
VERSION  
KEYWORDS EST.  
SOURCE Pneumocystis carinii f. sp. carinii.  
ORGANISM Pneumocystis carinii f. sp. carinii.  
Pneumocystis carinii f. sp. carinii.  
Eukaryota; Fungi; Fungi incertae sedis; Pneumocystidaceae; Pneumocystis.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman,J.C., Kovacs,J. and Cushion,M.  
TITLE Expressed sequence tags from Pneumocystis carinii  
JOURNAL Unpublished (2000)  
COMMENT Contact: Staben C  
School of Biological Sciences  
University of Kentucky  
101 Morgan Building, University of Kentucky, Lexington, KY 40506-0225, USA  
Tel: 606 257 2161  
Fax: 606 257 1717  
Email: staben@pop.uky.edu.  
Location/Qualifiers

```

source
1..26
/organism="Pneumocystis carinii f. sp. carinii"
/db_xref="taxon:38081"
/clone_lib="AGS-1"
/lab_host="E. coli"
/notes="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dT priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/project/Pneumocystis/"
BASE COUNT 13 a 2 c 1 g 10 t
ORIGIN
Query Match 100.0%; Score 7; DB 89; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtgata 7
Db 15 TTTGATA 21

RESULT 10
D18735 27 bp mRNA EST 12-DEC-1995
LOCUS MUSG01797 Mouse 3'-directed Mus musculus domesticus cDNA clone
DEFINITION md2066 3', mRNA sequence.
ACCESSION D18735
VERSION D18735.1 GI:1100704
KEYWORDS EST.
SOURCE western European house mouse.
ORGANISM Mus musculus domesticus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1..(bases 1 to 27)
REFERENCE Kawamoto,S., Okubo,K., Yoshii,J., Katsuki,M. and Matsubara,K.
AUTHORS Analysis of gene expression in mouse embryogenesis by 3'-directed
TITLE cDNA sequencing
JOURNAL Unpublished (1995)
COMMENT Contact: Kawamoto,S., Okubo,K., Yoshii,J., Katsuki,M. and Matsubara
.K
Institute for Cellular and Molecular Biology
Osaka University
3-1 Yamada-oka, Suita, Osaka 565, Japan.
FEATURES
Location/Qualifiers
1..27
/organism="Mus musculus domesticus"
/strain="C57BL/6J"
/db_xref="taxon:10092"
/clone_lib="md2066"
/tissue_type="decidual tissue (day 6.5-8.5 of gestation)"
BASE COUNT 11 a 4 c 4 g 8 t
ORIGIN
Query Match 100.0%; Score 7; DB 140; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtgata 7
Db 15 TTTGATA 21

RESULT 11
A2328549 27 bp DNA GSS 29-SEP-2000
LOCUS 1M0052D24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0052D24 F, DNA sequence.
ACCESSION A2328549

```

---

```

VERSION A2328549.1 GI:10388388
SOURCE GSS.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1..(bases 1 to 27)
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
AUTHORS Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
.M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL plasmid inserts
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0052 row: D column: 24
Seq primer: CGTTGTAAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 27.
FEATURES
Location/Qualifiers
1..27
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0052D24"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, P-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gil4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 5 a 5 c 9 g 8 t
ORIGIN
Query Match 100.0%; Score 7; DB 170; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtgata 7
Db 4 TTTGATA 10

RESULT 12
A2335603 27 bp DNA GSS 29-SEP-2000
LOCUS AZ335603
DEFINITION 1M0065E13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0065E13 R, DNA sequence.

```

```

ACCESSION A2335603
VERSION A2335603.1 GI:10404082
KEYWORDS GSS.
SOURCE Mus musculus.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0065 row: E column: 13
Seq primer: CACACAGGAACACGCTATGACC
Class: Plasmid ends
High quality sequence stop: 27.
FEATURES
Location/Qualifiers
source
1..27
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC1M0065E13"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gil4732114[gblAF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 9 a 8 c 1 g 9 t
ORIGIN

Query Match 100.0% Score 7; DB 170; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
Db 8 TTTGATA 2

RESULT 13
AU007442 30 bp mRNA EST 31-JUL-1998
LOCUS AU007442 Schizosaccharomyces pombe late log phase cDNA
DEFINITION AU007442 Schizosaccharomyces pombe late log phase cDNA

ACCESSION A2335603
VERSION A2335603.1 GI:10404082
KEYWORDS GSS.
SOURCE Mus musculus.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0065 row: E column: 13
Seq primer: CACACAGGAACACGCTATGACC
Class: Plasmid ends
High quality sequence stop: 27.
FEATURES
Location/Qualifiers
source
1..27
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC1M0065E13"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gil4732114[gblAF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 9 a 8 c 1 g 9 t
ORIGIN

Query Match 100.0% Score 7; DB 170; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
Db 8 TTTGATA 2

RESULT 13
AU007442 30 bp mRNA EST 31-JUL-1998
LOCUS AU007442 Schizosaccharomyces pombe late log phase cDNA
DEFINITION AU007442 Schizosaccharomyces pombe late log phase cDNA

ACCESSION AU007442
VERSION AU007442.1 GI:3343900
KEYWORDS EST.
SOURCE Schizosaccharomyces pombe
ORGANISM Schizosaccharomyces pombe
fission yeast.
Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomycetes.
AUTHORS Moriyoshi,M. and Mita,K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces
pombe
JOURNAL Unpublished (1998)
COMMENT Contact: Mitsuaki Moriyoshi
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: moriyoshi@nirs.go.jp.
FEATURES
Location/Qualifiers
source
1..30
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc02033"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/notes="vector: M13mpl9; The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mpl9 DNA and the direction of DNA
sequences was not always from 5' to 3'. The cDNA data of
Schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT 13 a 0 c 2 g 15 t
ORIGIN

Query Match 100.0% Score 7; DB 28; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtgata 7
Db 21 TTTGATA 27

RESULT 14
BE911076 30 bp mRNA EST 29-SEP-2000
LOCUS BE911076 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962130 5',
DEFINITION BE911076 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962130 5',
mRNA sequence.
ACCESSION BE911076
VERSION BE911076.1 GI:10408060
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 30)
NTH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9128 row: f column: 19
High quality sequence stop: 30.

```



FEATURES  
source

## Location/Qualifiers

1. .30  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_lib="IMAGE:3962130"  
/clone\_lib="NCI CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="3 months, virgin"  
/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 12 a 1 c 4 g 13 t

## ORIGIN

Query Match 100.0%; Score 7; DB 137; Length 30;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7

|||||

Db 16 TTTGATA 22

## RESULT 15

C20899/c

## LOCUS

C20899 30 bp mRNA EST 23-OCT-1996  
HUMGS0004973 Human adult (K.Okubo) Homo sapiens CDNA 3', mRNA  
sequence.

## ACCESSION

C20899

## VERSION

C20899.1 GI:1622009

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 30)

Okubo, K.

BodyMap: human gene expression database

Unpublished (1995)

Contact: Okubo, K.

Institute for Molecular and Cellular Biol

Osaka University

1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan

Tel: 06-877-5111(ex.3315)

Email: kousaku@imcb.osaka-u.ac.jp

Human Gene Signature. 3'-directed cDNA sequence. We are not

submitting the same cDNA sequence redundantly to DBJ since 1993.

For the abundance information of clones with this sequence in this

library and as well as in other 3'-directed libraries, see

http://www.imcb.osaka-u.ac.jp/bodymap'. The sequences of the clones

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

represented by this GS sequences is also found there.

Search completed: March 27, 2001, 07:38:23  
Job time: 4586 sec

FEATURES  
source

## Location/Qualifiers

1. .30  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human adult (K.Okubo)"  
/dev\_stage="adult"

BASE COUNT 14 a 6 c 5 g 5 t

## ORIGIN

Query Match 100.0%; Score 7; DB 139; Length 30;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtgata 7

|||||

Db 21 TTTGATA 15



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:17:04 ; Search time 2286.42 Seconds  
(without alignments)  
49.243 Million cell updates/sec

Title: US-09-380-826A-6  
Perfect score: 22  
Sequence: 1 ttttgannnnnnnnnnttgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_ph:\*
- 6: gb\_pl1:\*
- 7: gb\_pl2:\*
- 8: gb\_pr1:\*
- 9: gb\_pr2:\*
- 10: gb\_pr3:\*
- 11: gb\_ro:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: em\_fun:\*
- 15: em\_hum1:\*
- 16: em\_hum2:\*
- 17: em\_in:\*
- 18: em\_om:\*
- 19: em\_or:\*
- 20: em\_ov:\*
- 21: em\_pat:\*
- 22: em\_ph:\*
- 23: em\_pl:\*
- 24: em\_ro:\*
- 25: em\_sts:\*
- 26: em\_sy:\*
- 27: em\_un:\*
- 28: em\_v1:\*
- 29: gb\_ba3:\*
- 30: gb\_in1:\*
- 31: gb\_in2:\*
- 32: gb\_in3:\*
- 33: gb\_pl3:\*
- 34: gb\_pr4:\*
- 35: em\_ba1:\*
- 36: em\_ba2:\*
- 37: em\_htg1:\*
- 38: em\_htg2:\*
- 39: em\_htg3:\*
- 40: em\_htg4:\*
- 41: em\_htg5:\*
- 42: em\_htg6:\*
- 43: em\_htg7:\*

- 44: em\_htg8:\*
- 45: em\_htg9:\*
- 46: em\_htg10:\*
- 47: em\_hum3:\*
- 48: em\_hum4:\*
- 49: em\_hum5:\*
- 50: em\_hum6:\*
- 51: gb\_pr5:\*
- 52: gb\_pr6:\*
- 53: gb\_pr7:\*
- 54: gb\_htg1:\*
- 55: gb\_htg2:\*
- 56: gb\_htg3:\*
- 57: gb\_htg4:\*
- 58: gb\_htg5:\*
- 59: gb\_htg6:\*
- 60: gb\_htg7:\*
- 61: gb\_htg8:\*
- 62: gb\_htg9:\*
- 63: gb\_htg10:\*
- 64: gb\_htg11:\*
- 65: gb\_htg12:\*
- 66: gb\_htg13:\*
- 67: gb\_htg14:\*
- 68: gb\_htg15:\*
- 69: gb\_htg16:\*
- 70: gb\_htg17:\*
- 71: gb\_htg18:\*
- 72: gb\_htg19:\*
- 73: gb\_htg20:\*
- 74: gb\_htg21:\*
- 75: gb\_htg22:\*
- 76: gb\_htg23:\*
- 77: gb\_sts1:\*
- 78: gb\_sts2:\*
- 79: gb\_vil:\*
- 80: gb\_v12:\*
- 81: gb\_pat1:\*
- 82: gb\_pat2:\*
- 83: em\_htg0:\*
- 84: gb\_htg24:\*
- 85: gb\_pr8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	63.6	239	53	HSSSG03
2	14	63.6	435	78	HUM4STS161
3	14	63.6	668	77	G50458
4	14	63.6	1015	80	SRSSRSV
5	14	63.6	1257	3	BTAJ9225
6	14	63.6	1257	52	HSAJ9229
7	14	63.6	1258	3	BTAJ9224
8	14	63.6	1276	3	BTAJ9223
9	14	63.6	1287	53	HSU52077
10	14	63.6	1481	2	LFU60594
11	14	63.6	2321	32	DDU66913
12	14	63.6	2683	29	SPGROELGN
13	14	63.6	3666	81	A45819
14	14	63.6	3666	81	A45820
15	14	63.6	3915	33	SCYJL046W
16	14	63.6	4197	81	A45813
17	14	63.6	4197	81	A45814
18	14	63.6	4684	33	SCYJL045W
19	14	63.6	5620	11	MMTHREC02
20	14	63.6	10948	58	AC014767
21	14	63.6	11949	51	AF000145
					AF051649 Homo sapi
					M95263 Human chrom
					G50458 SHGC-79916
					D38547 Small round
					AJ009225 Bos tauru
					AJ009229 Homo sapi
					AJ009224 Bos tauru
					AJ009223 Bos tauru
					U52077 Human marin
					U60594 leptospira
					X89236 S.pyogenes
					A45819 Sequence 13
					A45820 Sequence 14
					Z43321 S.cerevisia
					A45813 Sequence 7
					A45814 Sequence 8
					Z43320 S.cerevisia
					U36757 Mus musculu
					AC014767 Drosophil
					AF000145 Homo sapi

```

c 22 14 63.6 12437 1 AB002102 Ureaplasma
c 23 14 63.6 12973 57 AC013026 Drosophila
c 24 14 63.6 12978 30 AC006805 Caenorhabd
c 25 14 63.6 28923 31 CEM03B6
c 26 14 63.6 37923 31 CELK07H8
c 27 14 63.6 42042 31 CEC34B4
c 28 14 63.6 53012 52 HSABLGR2
c 29 14 63.6 63156 65 AC026482 Homo sapi
c 30 14 63.6 63739 34 AF271897 Homo sapi
c 31 14 63.6 64064 71 AC083757 Homo sapi
c 32 14 63.6 64577 30 AE002828 Drosophila
c 33 14 63.6 64789 72 AC083839 Homo sapi
c 34 14 63.6 67987 58 AC016064 Homo sapi
c 35 14 63.6 68082 58 AC014851 Drosophila
c 36 14 63.6 71842 66 AC036189 Homo sapi
c 37 14 63.6 74188 61 AC021315 Homo sapi
c 38 14 63.6 74562 52 HS25D22 Human DNA s
c 39 14 63.6 78877 9 AC008408 Homo sapi
c 40 14 63.6 85127 64 AC025978 Homo sapi
c 41 14 63.6 86113 34 AF285442 Homo sapi
c 42 14 63.6 88735 72 AF165178 Homo sapi
c 43 14 63.6 88800 51 HS127M18 Homo sapi
c 44 14 63.6 8958 9 AC008967 Homo sapi
c 45 14 63.6 91470 6 AC011438 Genomic s

```

## ALIGNMENTS

```

RESULT 1
HSSG03 239 bp DNA 14-MAR-1999
LOCUS Homo sapiens squalene synthase gene, Intron 2, 5' end.
DEFINITION AF051649
ACCESSION AF051649.1 GI:4415976
VERSION 3 of 14
KEYWORDS human.
SEGMENT
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 239)
Jiang, G., McKenzle, T.L., Conrad, D.G. and Shechter, I.
JOURNAL Transcriptional regulation by lovastatin and 25-hydroxycholesterol
MEDLINE in HepG2 cells and molecular cloning and expression of the cDNA for
REFERENCE the human hepatic squalene synthase
AUTHORS J. Biol. Chem. 268 (17), 12818-12824 (1993)
TITLE 93286128
JOURNAL Direct Submission
MEDLINE Guan, G., Dai, P.H. and Shechter, I.
REFERENCE 2 (bases 1 to 239)
AUTHORS Direct Submission
TITLE Submitted (02-MAR-1998) Biochemistry and Molecular Biology,
JOURNAL Uniformed Services University of the Health Sciences, 4301 Jones
Bridge Road, Bethesda, MD 20814-4799, USA

```

## FEATURES

```

Source
1..239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8p22-p23.1"
33..>239
/gene="squalene synthase"
/number=2
43 a 61 c 70 g 65 t
BASE COUNT
ORIGIN

```

```

Query Match 63.6%; Score 14; DB 53; Length 239;
Best Local Similarity 63.6%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 tgttggannnnnnntttgata 22
|||||

```

```

Db 92 TGTTCGAAGCTACCTTTTGATA 113
RESULT 2
HUMASTS161
LOCUS HUM4STS161
DEFINITION Human chromosome 4 sequence-tagged site STS4-161, sequence tagged
site.
ACCESSION M95263
VERSION M95263.1 GI:177262
KEYWORDS STS; human chromosome 4; sequence tagged site.
SOURCE Homo sapiens, clone C4-177 from Los Alamos National Laboratory
chromosome 4 cosmid library, plate 4-1-2R, pos3C; vector SuperCos 1
(Stratagene).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 435)
Dugaiczky, A., Smith, K.A., Cox, D.R., Masters, S.B. and Myers, R.M.
TITLE The development of sequence-tagged sites for human chromosome 4
JOURNAL Hum. Mol. Genet. 2 (8), 1271-1288 (1993)
MEDLINE 94004872
COMMENT PCR components: 25 ng of human genomic DNA, 10 pmol of each
oligonucleotide, 200 micro-M dNTPs, 0.25 U Taq polymerase (Cetus)
in 10 micro-l of 50 mM KCl-20 mM Tris-HCl, pH 8.3 (at room temp),
2.5 mM MgCl-2. Initial denaturation at 94degC for 1.5 min, then 30
cycles of 94degC for 15 sec, 62degC for 23 sec, and 72degC for 30
sec, followed by a final extension at 72degC for 3.5 min, using a
Perkin-Elmer 9600 thermocycler. PCR-amplified product size 279 bp.
Sequence submitted by:
Human Genome Mapping Center
Box 0925
University of California San Francisco
San Francisco, CA 94143-0925 USA
Phone: (415) 502-1612 Fax: (415) 476-8391
e-mail: hgmcpbesecgl.ucsf.edu.
FEATURES
Source
1..435
/organism="Homo sapiens"
/db_xref="taxon:9606"
primer_bind complement(62..84)
primer_bind 316..340
BASE COUNT 128 a 64 c 75 g 162 t 6 others
ORIGIN
Query Match 63.6%; Score 14; DB 78; Length 435;
Best Local Similarity 63.6%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 tgttggannnnnnntttgata 22
|||||
Db 262 TGTTCGAATTCTGCTTTTGATA 283
RESULT 3
G50458
LOCUS SHGC-79916 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION G50458
ACCESSION G50458
VERSION G50458.1 GI:5221635
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 668)
Olivier, M. and Cox, D.R.
TITLE Unpublished, Olivier, M., Cox, D.R. (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Michael Olivier, David R. Cox

```

Stanford Human Genome Center  
 Stanford University School of Medicine  
 4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA  
 Tel: (650) 320-5800  
 Fax: (650) 320-5801  
 Email: olivier@shgc.stanford.edu  
 Primer A: GCACATGAGCTTTGGATTGCTT  
 Primer B: TGCCTTGTGCTTTAGCAATTTT  
 STS size: 278  
 PCR Profile:

Initial incubation: 95 degrees C for 10 minutes  
 Denaturation: 94 degrees C for 30 seconds  
 Annealing: 60 degrees C for 30 seconds  
 Polymerization: 72 degrees C for 23 seconds  
 PCR Cycles: 30  
 Thermal Cycler: Perkin Elmer 9700

Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 AmpliTaq Gold Polymerase: 0.07 units/ul  
 Total Vol: 5 ul

Buffer:  
 MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

Location/Qualifiers

#### FEATURES

source

1..668

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="8"  
 /clone\_lib="Human"

31..308

31..53

primer\_bind complement(286..308)  
 primer\_bind 104 c 113 g 221 t

#### BASE COUNT

230 a 104 c 113 g 221 t

#### ORIGIN

Query Match 63.6%; Score 14; DB 77; Length 668;  
 Best Local Similarity 63.6%; Pred. No. 1.2e+03;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22

|||||

Db 543 TGTGGAAAATTACCTTGATA 564

#### RESULT 4

SRSSRSV/c

LOCUS

SRSSRSV 1015 bp RNA VRL 08-FEB-1999  
 Small round structured virus genomic RNA, 3' terminal sequence containing ORF2 and ORF3.

ACCESSION

D38547

VERSION

D38547.1 GI:560054

KEYWORDS

SOURCE

Small round structured virus (isolate:patient; SRSV-CHIBA-407/87/J)  
 cDNA to genomic RNA.

ORGANISM

Utagawa, E.T.

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (17-OCT-1994) to the DDBJ/EMBL/GenBank databases. Etsuko T. Utagawa, National Institute of Health, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162, Japan (Tel:03-5285-1111(ex.2562), Fax:03-5285-1177)

#### REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

1..1015

/organism="small round structured virus"

/isolate="patient; SRSV-CHIBA-407/87/J"

/db\_xref="taxon:37141"

/note="SRSVs are currently recognised as a potential human calicivirus."

<1..298

/note="ORF2"

/codon\_start=2

/protein\_id="BAA07549.1"

/db\_xref="GI:1232112"

/translation="VPCLLPQBYITHFISEQAPIQGEAALLHYVDPDTRNRLNGEFLY

PGYLITCVNPSSSTGPQLPLDGVFASWSRFYQKPVGTAGPARGLRVRR"

298..924

/note="ORF3"

/codon\_start=1

/protein\_id="BAA07550.1"

/db\_xref="GI:1232113"

/translation="MAQAIIGAIAASAAGSALGAGIQAQAEAAALQAOYQODLTLOON

SFNHDKEMLYQYEMENSKLAKNLNRYSLQAGLSSDAARAVAGAPVTRLVDMGG

VRVAAPQSSATLTRSGNFMAVPLPAQPKOKPLASEGYNPAIDPVQRTASWVSQNS

RSWSPYHQALQTVVTPGSSSSSVSTAPRGYFNTDLPLFLANLR"

BASE COUNT 229 a 251 c 233 g 302 t

ORIGIN

Query Match 63.6%; Score 14; DB 80; Length 1015;  
 Best Local Similarity 63.6%; Pred. No. 1.1e+03;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22

|||||

Db 481 TGTGGACATTTCCATTTGATA 460

#### RESULT 5

BTJ9225

LOCUS

BTJ9225 1257 bp DNA MAM 11-JUL-2000

Bos taurus mariner related transposon Hsmar1, clone btmlc4.

ACCESSION

AJ009225

VERSION

AJ009225.1 GI:9187451

KEYWORDS

transposon.

SOURCE

Bos taurus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 1257)

Demattell, M.V., Auge-Gouillou, C., Pollet, N., Meunier-Rotival, M. and

Bigot, Y.

Features of the mammal mar1 transposons in the human, sheep, cow

and mouse genomes and implications for their evolution

Mamm. Genome In press

2 (bases 1 to 1257)

Bigot, Y.

Direct Submission

TITLE

Submitted (04-JUN-1998) Bigot Y., Institut de Recherche sur la

Biologie de l'Insectes, UPRESA CNRS 6035, Facult des Sciences, Parc

Grandmont, 37200 Tours, FRANCE

Grandmont, 37200 Tours, FRANCE

Location/Qualifiers

1..1257

/organism="Bos taurus"

/transposon="Hsmar1"

/db\_xref="taxon:9913"

/clone="btmlc4"

BASE COUNT 388 a 263 c 263 g 343 t

Wed Mar 28 14:03:23 2001

```

ORIGIN
  Query Match      63.6%; Score 14; DB 3; Length 1257;
  Best Local Similarity 63.6%; Pred. No. 1.1e+03;
  Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22
    |||||||
Db 36 TGTGGAAATTTGCCATTGATA 57

RESULT 6
HSAJ9229      1257 bp DNA PRI 11-JUL-2000
LOCUS Homo sapiens mariner related transposon Hsmar1, clone mlehsc3.
DEFINITION AJ009229
ACCESSION AJ009229.1 GI:9187522
VERSION transposon.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1257)
AUTHORS Demattei,M.V., Auge-Gouillou,C., Pollet,N., Meunier-Rotival,M. and
Bigot,Y.
TITLE Features of the mammal marl transposons in the human, sheep, cow
and mouse genomes and implications for their evolution
JOURNAL Mamm. Genome In press
REFERENCE 2 (bases 1 to 1257)
AUTHORS Bigot,Y.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-1998) Bigot Y., Institut de Recherche sur la
Biologie de l'Insects, UPRESA CNRS 6035, Facult des Sciences, Parc
Grandmont, 37200 Tours, FRANCE
FEATURES
  source 1. .1257
  /organism="Homo sapiens"
  /transposon="Hsmar1"
  /db_xref="taxon:9606"
  /clone="mlehsc3"

BASE COUNT 381 a 263 c 264 g 349 t
ORIGIN

  Query Match      63.6%; Score 14; DB 52; Length 1257;
  Best Local Similarity 63.6%; Pred. No. 1.1e+03;
  Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22
    |||||||
Db 36 TGTGGAAATTTGCCATTGATA 57

RESULT 7
HSAJ9224      1258 bp DNA MAM 11-JUL-2000
LOCUS Bos taurus mariner related transposon Hsmar1, clone btmlec3.
DEFINITION AJ009224
ACCESSION AJ009224.1 GI:9187450
VERSION transposon.
KEYWORDS Bos taurus.
SOURCE Bos taurus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1258)
AUTHORS Demattei,M.V., Auge-Gouillou,C., Pollet,N., Meunier-Rotival,M. and
Bigot,Y.
TITLE Features of the mammal marl transposons in the human, sheep, cow
and mouse genomes and implications for their evolution
JOURNAL Mamm. Genome In press
REFERENCE 2 (bases 1 to 1258)

BASE COUNT 388 a 261 c 248 g 379 t
ORIGIN

  Query Match      63.6%; Score 14; DB 3; Length 1276;
  Best Local Similarity 63.6%; Pred. No. 1.1e+03;
  Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22
    |||||||
Db 37 TGTGGAAATTTGCCATTGATA 58

RESULT 9
HSAJ92077      1287 bp DNA PRI 26-JAN-1998
LOCUS Human mariner1 transposase gene, complete consensus sequence.
DEFINITION U52077
ACCESSION U52077.1 GI:1263080
VERSION U52077.1 GI:1263080

```

```

KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 1287)
            Robertson,H.M. and Zumpano,K.L.
            Molecular evolution of an ancient mariner transposon, Hsmar1, in
            the human genome
JOURNAL     Gene 205 (1-2), 203-217 (1997)
MEDLINE     98121293
REFERENCE   2 (bases 1 to 1287)
AUTHORS     Robertson,H.M.
TITLE       Direct Submission
JOURNAL     Submitted (21-MAR-1996) Hugh M. Robertson, Entomology, University
            of Illinois at Urbana-Champaign, 505 S. Goodwin, Urbana, IL 61801,
            USA
FEATURES
Source      Location/Qualifiers
            1..1287
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /transposon="Hsmar1"
            /note="consensus sequence based on 20 unique sequences"
repeat_region 1..30
            /rpt_type=Inverted
            179..1210
CDS         /codon_start=1
            /product="mariner transposase"
            /protein_id="AAC52010.1"
            /db_xref="GI:1263081"
            /translation="MEMLDKKQIRALFLFEKMGKRAETTRNNNAGPGCGTANERT
            VWMFKCKGDELSRQSEVNDQKAIIEADPLTTTREVAEELNVDHSTV
            VHLKQIKGVKLDKWPHELSQNRKRFVSSLIILRNNEPFLDRIVTCDEKWL
            YNRRPAQWLDREAPKHFPPKPNLHQKVMVYVWWSAAGLIHYSFLNPGETITSEKY
            AQCIDEMHKLQRLQPALVNRKGPILLHDNAPHAQPTLQKLNELGYEVLPHYPSP
            DISPDYHFEKHLDFLQGRPHNQDAENAFQEVESRSTDFYATGINKLISRQKC
            VDCNGSYFD"
repeat_region 1256..1287
BASE COUNT 381 a 284 c 278 g 344 t
ORIGIN
            1..1287
            |||||
            36 TGTGGAAATTGCCGTTTGATA 57

Query Match      63.6%; Score 14; DB 53; Length 1287;
Best Local Similarity 63.6%; Pred. No. 1.le+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttgtggannnnnnnttgata 22
|||||
Db 36 TGTGGAAATTGCCGTTTGATA 57

RESULT 10
LOCUS      LFU60594
DEFINITION Leptospiira fainei 16S ribosomal RNA gene, partial sequence.
ACCESSION  U60594
VERSION     U60594.1 GI:1408219
KEYWORDS   .
SOURCE     Leptospiira fainei.
ORGANISM   Bacteria; Spirochaetales; Leptospiaceae; Leptospiira.
REFERENCE  1 (bases 1 to 1481)
AUTHORS    Perolat,P., Chappal,R.J., Adler,B., Baranton,G., Bulach,D.M.,
            Billinghamurst,M.L., Letocart,M., Merien,F. and Serrano,M.S.
            Leptospiira fainei sp. nov., isolated from pigs in Australia
            Int.J.Syst. Bacteriol. 48 Pt 3, 851-858 (1998)
TITLE      98404550
MEDLINE
REFERENCE  2 (bases 1 to 1481)
AUTHORS    Adler,B., Chappal,R.J., Baranton,G., Bulach,D.M.,
            Billinghamurst,M.L., Letocart,M., Merien,F., Serrano,M.S. and
            Perolat,P.
TITLE      Direct Submission

JOURNAL     Submitted (12-JUN-1996) Microbiology, Monash University, Wellington
            Rd., Clayton, VIC 3168, Australia
FEATURES
Source      Location/Qualifiers
            1..1481
            /organism="Leptospiira fainei"
            /strain="Hurstbridge"
            /db_xref="taxon:48782"
            <1..>1481
            /product="16S ribosomal RNA"
BASE COUNT 391 a 335 c 439 g 314 t
ORIGIN
            1..1481
            |||||
            155 TGTGGATCACAGATTTGATA 176

Query Match      63.6%; Score 14; DB 2; Length 1481;
Best Local Similarity 63.6%; Pred. No. 1.le+03;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttgtggannnnnnnttgata 22
|||||
Db 155 TGTGGATCACAGATTTGATA 176

RESULT 11
LOCUS      DDU66913/c
DEFINITION Dictyostellium discoideum ORF DG1040 gene, partial cds.
ACCESSION  U66913
VERSION     U66913.1 GI:1519537
KEYWORDS   .
SOURCE     Dictyostellium discoideum.
ORGANISM   Dictyostellium discoideum
REFERENCE   1 (bases 1 to 2321)
AUTHORS     Loomis,W.F.
TITLE       Direct Submission
JOURNAL     Submitted (15-AUG-1996) Dept. of Biology 0322, University of
            California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES
Source      Location/Qualifiers
            1..2321
            /organism="Dictyostellium discoideum"
            /strain="AX4"
            /db_xref="taxon:44689"
            /cell_line="AX4"
            join(370..732,813..>2321)
            /note="ORF DG1040"
            <370..732
            /number=1
            join(370..732,813..>2321)
            /note="ORF DG1040; initially derived from a plasmid
            disrupted gene"
            /codon_start=1
            /protein_id="AAB07544.1"
            /db_xref="GI:1519538"
            /translation="MSTKPEIDLLKSEVNDSINNDDNNNSINNDDNNNNKNEE
            TSEYDPPNIFNFKLSSDKSPYKWTWLOVGSNNNEQTDNSTINLSEVATQIKN
            LISLTKKYSKNSSELQINQYQSAEIIYIRCLVSDISFNKNTWPIITTSITSA
            TSTTPTPTPTTASSTAASSSSNLQIKLQDFESRLTKQPHFISVLCKAFDGYE
            VLNNEFLHFSSTLTPSHLMLGSLQSLDKVEQADKFLNLTILQSLQNTKS
            LPENLLHOLLYFKIQTNOLOQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQ
            QNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNN
            QNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNN
            SILSKVQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNN
            TNNNNNNNNNDENKIPYITFNDLNKETSNTSPSSSPKQSENNSSSTTTT
            NNNNNNNNTNIPDENISIFVDVIKELYITDWDIVIREMDCPIVNNNDQGLAFI
            LAVYKATQVQQQKPPFIDFLDRVWVNSLSIQIFLKIAIQSDFFPFQNSTKKRID
            "
            813..>2321
            /number=2
            2124
            /note="site of plasmid insertion in mutant organisms;
            mutants form crinkled, branching fingers; neither spores
            nor stalk cells were seen"
BASE COUNT 1055 a 338 c 210 g 718 t
ORIGIN
            1..2321
            |||||
            36 TGTGGAAATTGCCGTTTGATA 57

```

Query Match 63.6%; Score 14; DB 32; Length 2321;  
 Best Local Similarity 63.6%; Pred. No. 1.le+03;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22  
 |||||  
 Db 2069 TGTGGATATAATCTTGTGATA 2048

RESULT 12  
 LOCUS SPGROELGN 2683 bp DNA BCT 26-SEP-1997  
 DEFINITION S. Pyogenes DNA for groEL gene.  
 ACCESSION X89236  
 VERSION X89236.1 GI:2462691  
 KEYWORDS groEL gene; heat shock protein 60 (GroEL) like protein.  
 SOURCE Streptococcus pyogenes.  
 ORGANISM Streptococcus pyogenes  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

REFERENCE 1 (bases 1 to 2683)  
 AUTHORS Podbielski, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JUN-1995) A. Podbielski, Institut f Med  
 Mikrobiologie, an der Rhein Westf. Techn. Hochschule, Aachen,  
 Pauwelsstr Klinikum, 52057 Aachen, FRG  
 2 (bases 1 to 2683)  
 Pohl, B., Podbielski, A. and Zarges, I.  
 AUTHORS Unpublished  
 JOURNAL Related sequences M81132, M84965.  
 COMMENT

FEATURES  
 Source  
 1..2683  
 Location/Qualifiers  
 /organism="Streptococcus pyogenes"  
 /strain="serotype M49"  
 /isolate="CS101"  
 /db\_xref="taxon:1314"  
 1..1446  
 /gene="groEL"  
 <1..1446  
 /gene="groEL"  
 /codon\_start=1  
 /transl\_table=11  
 /product="heat shock protein"  
 /protein\_id="CAA61520.1"  
 /db\_xref="GI:2462692"  
 /db\_xref="SPTREMBL:O33733"  
 /translation="DHFFENMGAKLYSEVASKTNDIAGDGTATVLTQAIIVHEGLKNV  
 TGANPIGIRGIETATATAEALKAVAQPVSGKEAIAQVAAYVSRSEKVGGEYISEAM  
 ERVNDGVITIEESRGMETELVVEGMDQFDRGYLSQYMTDNERKMWADLENPFILIMD  
 KRVSNIQDILPLEELVLTNRPLLIADVDGEALPTLVLNKRITFNVAVKAPFG  
 DRKAMLEDIALITGCTVITEDGLELKDATMALGQAAKITVDKDSIVVEGSGSE  
 ATANRALIKSOLETTSDDFREKLEQLAKLAGGAVIKVGAPTEALKEMKLRID  
 ALNAPRAVEEIVAGGCTALTIVTEKVALEGGDATGRNIVLRALKEEPPVROIALN  
 AGESVVIDKLKNSPAGTGFGNAATGEWMDMKTGIIDPVKVTRSALQNAASVASLIL  
 TTEAVYANKPEAPAPAMPAGMDPGMMGGF"  
 1509..1564  
 repeat\_region  
 /rpt\_type=INVERTED  
 BASE COUNT 808 a 490 c 593 g 792 t  
 ORIGIN

Query Match 63.6%; Score 14; DB 29; Length 2683;  
 Best Local Similarity 63.6%; Pred. No. 1.le+03;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttttgannnnnnnttgata 22  
 |||||  
 Db 963 TGTGGAGCTCCTACTTTGTATA 942

RESULT 13

A45819/c  
 LOCUS A45819 3666 bp DNA PAT 07-MAR-1997  
 DEFINITION Sequence 13 from Patent WO9519371.  
 ACCESSION A45819  
 VERSION A45819.1 GI:2300192  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 3666)  
 AUTHORS Deblaere, R. Y., Desomer, J. and Dhaese, P.  
 TITLE EXPRESSION OF SURFACE LAYER PROTEINS  
 JOURNAL Patent: WO 9519371-A 13 20-JUL-1995;  
 SOLVAY (BE)  
 FEATURES  
 Location/Qualifiers  
 1..3666  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 1281 a 662 c 676 g 1047 t  
 ORIGIN

Query Match 63.6%; Score 14; DB 81; Length 3666;  
 Best Local Similarity 63.6%; Pred. No. 1.le+03;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 Qy 1 ttttgannnnnnnttgata 22  
 |||||  
 Db 2652 TCTTGAGCAACTGGTTTGATA 2631

RESULT 14  
 LOCUS A45820/c 3666 bp DNA PAT 07-MAR-1997  
 DEFINITION Sequence 14 from Patent WO9519371.  
 ACCESSION A45820  
 VERSION A45820.1 GI:2300193  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 3666)  
 AUTHORS Deblaere, R. Y., Desomer, J. and Dhaese, P.  
 TITLE EXPRESSION OF SURFACE LAYER PROTEINS  
 JOURNAL Patent: WO 9519371-A 14 20-JUL-1995;  
 SOLVAY (BE)  
 FEATURES  
 Location/Qualifiers  
 1..3666  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 1..3666

/note="unnamed protein product; Protein sequence is in  
 conflict with the conceptual translation"  
 /codon\_start=1  
 /protein\_id="CAA02849.1"  
 /db\_xref="GI:2300194"  
 /translation="AQVNDYKISGYAKEAVALVDQGVIOGDTNGNPNPLNTVTRAQ  
 AAEIFTKALELEANGDNFKDKAGAWTINSIAAVANGIFEGVSATFEPNKLSTRS  
 EAAKILVEAFGLEGEADLSEFADASQVPAKRYLEIAVANGIFEGTDANKLPNNSI  
 TRQDFALVKRFTVDKVEGETPEEAFAVKAINNTTVEVTFEEETNVQALNFKIEGLEI  
 KNASVQTKNKKVVLTTTEAQTADKEYVLTLDTGETIGGFGVAAVPTKVELVSSAVQY  
 KLGOEVKVOAKVTVAEGOSKAGIPVTETPGNNNDGVVPTLTGALTNEEGIATYSY  
 RYKGTDEVTAYATGDRSKFSLSGYFHWGVDILSVEEVTTCASVNGANKYKVTYKN  
 PKTGPEANKTFNVGVFNMMNVTSDKVANKYVANGVAKQLQSLNGTALDRAQITTSKGE  
 ATFTVSGTNAATPVVYDLHSTNNSTSNKYSASALQTTASKVTFALQAQETIELTR  
 ADNAGEVAAGTATNGREYKVIKDKAGNLAKNEIVNFAFNEDKDRVISTVNAKPVDT  
 DPDTAVYFTGDKAKOISVTKNDKGEATIGSVTVDYATPIAMIDINTSDAKQGLD  
 EGEKAVAPISYFOAPYLDGSAIKAYKKSDLNKAVTKFDGSETVAFAELVNSQSKV  
 TGTSTKATYTYNTGANDIKVNOVISPNRSYTVTYTSSGTCTVITPAKNLEVT  
 VDGKTTAVKVIATGIANTDGDYAKTAKATATATATNATNPNSTVGTATQKFNADSC  
 SNSNSIWFAKNPNVAYSGKTYKFGANGNEVEGEAREALLIQTAYEQGKVTISY  
 NYVDGTVTFKVISAVNSSTEAIKPVAPTTTTPAATGALTLPAGGLVDLTTATNTIG  
 ISLADADLNVSATTVDTATVSLKDSANNSLSLTTLVETGANTGVFATTVOAGTLLSLTA



GTLTVTYADAKNAAGVAENITASVTLKTKTGTGTAITSDDTFTQGVLPNSATAAEYTSKSA  
ITTAITGAGGFLINDNAGQVINLAGKGAQGVADAINATFAGTATVSGDKVVKRS  
TTGVSEVEVTFSSVQVNLNAVYNGKQOVVAGTAAATKAFITITLTSVGKVVIDGVEY  
TAVAFCTAPTANTFVVESSAANTLASVADQAAANLAATIDTLNTADKFTASATGATITLT  
STVTPVGTITTEPVTLK"

BASE COUNT 1281 a 662 c 676 g 1047 t  
ORIGIN

Query Match 63.6%; Score 14; DB 81; Length 3666;

Best Local Similarity 63.6%; Pred. No. 1e+03; 8; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22

||||| |||||

Db 2652 TGTGGAGCAACTGGTTTGATA 2631

RESULT 15

SCYJL046W

LOCUS SCYJL046W 3915 bp DNA PLN 11-AUG-1997

DEFINITION S.cerevisiae chromosome X reading frame ORF YJL046w.

ACCESSION 249321 Y13136

VERSION 249321.1 GI:1008176

KEYWORDS

SOURCE

baker's yeast.

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;

Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 3915)

Pohl, T.M. and Aljinovic, G.

JOURNAL

Unpublished

REFERENCES

MIPS.

Direct Submission

Submitted (25-SEP-1995) Data collected by MIPS on behalf of the

European yeast chromosome X sequencing project. MIPS at the

Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

Martinsried, FRG; E-mail: Mewes@mips.emblnet.org

FEATURES

Source

CDS

1..3915  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
/chromosome="X"  
152..1507  
/note="ORF YJL046w"  
/codon\_start=1  
/protein\_id="CAA89337.1"  
/db\_xref="GI:1008177"  
/db\_xref="SWISS-PROT:P47051"  
/translation="MRSASSYLIRNIEHPKISELTYVEFLMNTKEENTRLFFDVYIMS  
MMLSNWALSPRYVQGRNLIHCTTLFHLTRWAKDADOKYHDINSYENMTSPNDNVS  
ILQDEGKSDYDTKASMEEDISAFNKDLYNFNIGYAKQIMSASOLENIVRAKGRFV  
IQSLSTSPYVNLALENYFKNTPRAKRPDNCRLLLYINDRCVAVIGKQNLNWOEVDLA  
KLKSNFELLRRFSGGTVLHLDGNVNSYLTLSREKFTKFNKMTIKWLSINPELR  
LDLNERGDIQDGKISQAYKIAGKAVHATMLLNADLEQFSGLLPELSPNNWEWE  
SSGVHSVKSKINVGIIITNQFIQVIAVSERFQMTFKVDGEIPIYICDEFKSIINDEKDA  
MNTLQSQWKVSPGKFSVKIKDKGLTIKVKRGMIVDCDRNDLIGLEFKFLENIDSY  
T"

trNA

2136..2224  
/gene="ty(GTA)JL - systematic name"

/note="tRNA-Tyr - common name; anticodon gene: GTA;

contains intron 14 nt"

gene

2136..2224  
/gene="ty(GTA)JL - systematic name"

complement(2445..2735)

/note="delta-remnant"

complement(2736..3067)

/note="solo delta"

trNA

3266..3337  
/gene="tr(TCT)JL2 - systematic name"

/note="tRNA-Arg2 - common name; anticodon gene: TCT"

gene

3266..3337

trNA  
/gene="tr(TCT)JL2 - systematic name"  
3348..3419  
/gene="td(GTC)JL2 - systematic name"  
/note="tRNA-Asp - common name; anticodon gene: GTC"  
3348..3419  
/gene="td(GTC)JL2 - systematic name"  
BASE COUNT 1258 a 667 c 698 g 1292 t  
ORIGIN

Query Match 63.6%; Score 14; DB 33; Length 3915;

Best Local Similarity 63.8%; Pred. No. 1e+03; 8; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22

||||| |||||

Db 2401 TGTGGAAATTCATTTTGTGATA 2422

Search completed: March 27, 2001, 08:17:06

Job time: 5903 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:22:59 ; Search time 207.51 Seconds  
(without alignments)  
39.827 Million cell updates/sec.

Title: US-09-380-826A-6

Perfect score: 22  
Sequence: 1 ttgtggannnnnnnnttgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
1: /cgn2\_2/gcgdata/geneseq/geneseq/NA1980.DAT:\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/NA1981.DAT:\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/NA1982.DAT:\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/NA1983.DAT:\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/NA1984.DAT:\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/NA1985.DAT:\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/NA1986.DAT:\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/NA1987.DAT:\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/NA1988.DAT:\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/NA1989.DAT:\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/NA1990.DAT:\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/NA1991.DAT:\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/NA1992.DAT:\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/NA1993.DAT:\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/NA1994.DAT:\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/NA1995.DAT:\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/NA1996.DAT:\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/NA1997.DAT:\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/NA1998.DAT:\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/NA1999.DAT:\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	63.6	22	V58901	Leptospira rRNA ge
2	14	63.6	22	V58897	L. fainei nucleoti
3	14	63.6	22	V58900	Leptospira nucleot
4	14	63.6	972	X95627	Nucleic acid seque
5	14	63.6	1477	V58896	L. fainei nucleoti
6	14	63.6	1661	X86155	DNA encoding a SLR
7	14	63.6	4197	Q99430	B. sphaericus STR
8	13	59.1	396	X34752	DNA encoding OMP-1
9	13	59.1	828	X216735	Human gene express
10	13	59.1	1079	X20676	Polynucleotide seq
11	13	59.1	1482	V18099	Nucleotide sequenc
12	13	59.1	1786	T60350	MAP kinase #2 codi

13	13	59.1	2190	20	X07102	Staphylococcus aur
14	13	59.1	2508	18	T67197	zebrafish retinoid
15	13	59.1	3519	21	Z94941	Human carbohydrate
16	13	59.1	3519	21	Z94948	Human carbohydrate
17	13	59.1	4012	18	V74358	Staphylococcus aur
18	13	59.1	5253	21	A26868	Essential Staphylo
19	13	59.1	9082	18	V74396	Staphylococcus aur
20	13	59.1	10723	18	T49304	CDNA encoding poly
21	13	59.1	10723	18	T49303	CDNA sequence enco
22	13	59.1	11802	18	V74381	Staphylococcus aur
23	13	59.1	15614	20	X12982	Enterococcus faeca
24	13	59.1	1230025	20	X91990	Nucleotide sequenc
25	13	59.1	1664976	19	V21209	Methanococcus jann
26	12.4	56.4	30	21	Z40032	Oligonucleotide fo
27	12.4	56.4	30	21	Z40033	Oligonucleotide fo
28	12.4	56.4	200	19	X12115	Human blaIIelic po
29	12.4	56.4	300	20	Z13959	Human gene express
30	12.4	56.4	300	20	Z12971	Human gene express
31	12.4	56.4	310	20	X20027	Enterococcus faeca
32	12.4	56.4	414	20	X20026	Enterococcus faeca
33	12.4	56.4	706	20	X20469	Human secreted pro
34	12.4	56.4	814	20	X35887	CDNA encoding a pr
35	12.4	56.4	877	18	V75289	Staphylococcus aur
36	12.4	56.4	915	19	V24146	Homo sapiens BARD1
37	12.4	56.4	963	21	A05842	Group B Streptococ
38	12.4	56.4	1034	18	T72785	Metastasis inducin
39	12.4	56.4	1131	18	V25036	H. pylori cellular
40	12.4	56.4	1152	18	V24747	H. pylori ORF hp4p
41	12.4	56.4	1248	21	A05803	Group B Streptococ
42	12.4	56.4	1249	19	X14031	H. pylori GHPO 886
43	12.4	56.4	1322	20	X13419	Enterococcus pneu
44	12.4	56.4	1353	18	X30790	Streptococcus pneu
45	12.4	56.4	1356	19	V04491	A. thaliana p-hydr

#### ALIGNMENTS

RESULT 1  
V58901  
ID V58901 standard; DNA; 22 BP.  
XX  
AC V58901;  
XX  
DT 20-JAN-1999 (first entry)  
XX  
DE Leptospira rRNA gene nucleotide sequence.  
XX  
KW Infection; pathogenic Leptospira; protective immunity; therapy;  
XX diagnosis; ss.  
XX  
OS Leptospira sp.  
XX  
PN WO9840099-A1.  
XX  
PD 17-SEP-1998.  
XX  
PF 06-MAR-1998; 98WO-AU00145.  
XX  
PR 07-MAR-1997; 97AU-0005494.  
XX  
(AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
PA (PIGR-) PIG RES & DEV CORP.  
PA  
PI Nucleic acid seque  
PI  
XX  
DR WPI; 1998-520791/44.  
XX  
PT New isolated pathogenic Leptospira bacterium - useful for, e.g  
PT developing products for conferring protective immunity, and for  
PT prophylactic or therapeutic treatment  
XX  
PS Claim 15; Page 72; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 CC L. fainei. The LS bacteria can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 63.6%; Score 14; DB 19; Length 22;  
 Best Local Similarity 63.6%; Pred. No. 33;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttgtggannnnnnntttgata 22  
 |||||  
 Db 1 ttgtggatcacagatttgata 22  
 |||||

RESULT 2  
 V58897  
 ID V58897 standard; DNA; 22 BP.  
 XX  
 AC V58897;  
 XX  
 DT 20-JAN-1999 (first entry)  
 XX  
 DE L. fainei nucleotide sequence.  
 DE  
 XX Infection; pathogenic Leptospira; protective immunity; therapy;  
 KW diagnosis; ss.  
 KW  
 XX Leptospira fainei.  
 OS  
 XX New isolated pathogenic Leptospira bacterium - useful for, e.g.  
 PN developing products for conferring protective immunity, and for  
 XX prophylactic or therapeutic treatment  
 PD 17-SEP-1998.  
 XX  
 PF 06-MAR-1998; 98WO-AU00145.  
 XX  
 PR 07-MAR-1997; 97AU-0005494.  
 XX  
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 PA (PIGR-) PIG RES & DEV CORP.  
 XX  
 PI Chappel RJ;  
 XX  
 DR WPI; 1998-520791/44.  
 XX  
 CC This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 CC L. fainei. The LS bacteria can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 63.6%; Score 14; DB 19; Length 22;  
 Best Local Similarity 63.6%; Pred. No. 33;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ttgtggannnnnnntttgata 22  
 |||||  
 Db 1 ttgtggatcacagatttgata 22  
 |||||

RESULT 3  
 V58900  
 ID V58900 standard; DNA; 22 BP.  
 XX  
 AC V58900;  
 XX  
 DT 20-JAN-1999 (first entry)  
 XX  
 DE Leptospira nucleotide sequence.  
 DE  
 XX Infection; pathogenic Leptospira; protective immunity; therapy;  
 KW diagnosis; ss.  
 KW  
 XX Leptospira sp.  
 OS  
 XX WO9840099-A1.  
 PN  
 XX 17-SEP-1998.  
 PD  
 XX 06-MAR-1998; 98WO-AU00145.  
 PF  
 XX 07-MAR-1997; 97AU-0005494.  
 PR  
 XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 PA (PIGR-) PIG RES & DEV CORP.  
 XX  
 PI Chappel RJ;  
 XX  
 DR WPI; 1998-520791/44.  
 XX  
 CC New isolated pathogenic Leptospira bacterium - useful for, e.g.  
 PT developing products for conferring protective immunity, and for  
 PT prophylactic or therapeutic treatment  
 PS Claim 15; Page 72; 94pp; English.

This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 CC L. fainei. The LS bacteria can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 SQ Sequence 22 BP; 3 A; 0 C; 4 G; 7 T; 8 other;

Query Match 63.6%; Score 14; DB 19; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtggannnnnnntttgata 22  
 |||||  
 Db 1 ttgtggannnnnnntttgata 22  
 |||||

RESULT 4  
 X99627/c  
 ID X99627 standard; DNA; 972 BP.  
 XX  
 AC X99627;  
 XX  
 DT 05-OCT-1999 (first entry)  
 DE  
 XX Nucleic acid sequence from U. urealyticum.  
 XX

KW Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;  
 KW human urogenital tract; pregnancy; neonatal disease; drug therapy;  
 XX suppurative arthritis; ss.  
 OS Ureaplasma urealyticum.  
 PN WO9939007-A1.  
 XX  
 XX  
 PD 05-AUG-1999.  
 XX  
 XX 29-JAN-1999; 99WO-US01972.  
 XX  
 XX 30-JAN-1998; 98US-0073189.  
 XX (UABR-) UAB RES FOUND.  
 PA  
 XX Cassell GH, Chen EY, Glass JT, Glass JS, Heiner CR;  
 PI Lefkowitz E;  
 XX WPI; 1999-469343/39.  
 DR  
 XX Detection of Ureaplasma urealyticum using novel genes, probes and  
 XX primers  
 XX  
 XX Claim 1; Page 82; 110pp; English.  
 XX The present invention provides methods for the detection and diagnosis  
 CC of Ureaplasma urealyticum infection. It provides novel genes (X99501-681)  
 CC that can be used as a source of primers and probes for the detection and/  
 CC or quantification of U. urealyticum in a biological sample. The probes  
 CC that can be used in the method of the invention by forming target:probe  
 CC complex is complementary to a region selected from one of the 181  
 CC nucleic acid sequences (X99501-681). U. urealyticum is an opportunistic  
 CC pathogen of the human urogenital tract that is a significant cause of  
 CC adverse pregnancy outcome, neonatal disease, and suppurative arthritis.  
 CC As the infections are commonly asymptomatic, it is important to have  
 CC specific and sensitive methods for detecting their presence in a patient.  
 CC Also, as the pathogen has no current antibiotic directed specifically  
 CC against it, it would be advantageous to isolate and detect gene sequences  
 CC which are unique to it, and utilise these as a basis for diagnosis of  
 CC U. urealyticum infection as well as to develop new and improved drug  
 CC therapies. The present invention provides such novel polynucleotide  
 CC sequences (X99501-681).  
 XX  
 XX Sequence 972 BP; 309 A; 127 C; 98 G; 438 T; 0 other;

Query Match 63.6%; Score 14; DB 20; Length 972;  
 Best Local Similarity 63.6%; Pred. No. 53;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 tgttgannnnnnntttgata 22  
 |||||  
 Db 931 TGTGGACTATTAGTTTGATA 910

RESULT 5  
 V58896  
 ID V58896 standard; DNA; 1477 BP.  
 XX  
 AC V58896;  
 XX  
 XX 20-JAN-1999 (first entry)  
 DT  
 XX L. fainei nucleotide sequence.  
 DE  
 XX Infection; pathogenic Leptospira; protective immunity; therapy;  
 KW diagnosis; ss.  
 KW Leptospira fainei.  
 XX  
 OS  
 XX WO9840099-A1.

PD 17-SEP-1998.  
 XX  
 PF 06-MAR-1998; 98WO-AU00145.  
 XX  
 PR 07-MAR-1997; 97AU-0005494.  
 XX  
 XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 PA (PIGR-) PIG RES & DEV CORP.  
 XX  
 XX Chappel RJ;  
 PI  
 XX WPI; 1998-520791/44.  
 DR  
 XX New isolated pathogenic Leptospira bacterium - useful for, e.g  
 PT developing products for conferring protective immunity, and for  
 PT prophylactic or therapeutic treatment  
 XX  
 XX Claim 15; Page 69-70; 94pp; English.  
 XX  
 CC This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar Hurstbridge or the species  
 CC L. fainei. The LS bacterium can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 XX Sequence 1477 BP; 390 A; 334 C; 439 G; 314 T; 0 other;

Query Match 63.6%; Score 14; DB 19; Length 1477;  
 Best Local Similarity 63.6%; Pred. No. 56;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 tgttgannnnnnntttgata 22  
 |||||  
 Db 154 tgttgatcacagatttgata 175

RESULT 6  
 X86155/C  
 ID X86155 standard; DNA; 1661 BP.  
 XX  
 AC X86155;  
 XX  
 XX 22-SEP-1999 (first entry)  
 DT  
 XX DNA encoding a Streptococcus pyogenes heat shock protein (Hsp)60-2.  
 DE  
 XX Heat shock protein; Hsp60-2; immune response; immunological carrier;  
 KW cancer control; tumour; sarcoma; cancer; gene therapy; ss.  
 KW  
 XX Streptococcus pyogenes.  
 OS  
 XX WO9935270-A1.  
 PN  
 XX  
 XX 15-JUL-1999.  
 PD  
 XX 29-DEC-1998; 98WO-CA01203.  
 PF  
 XX 31-DEC-1997; 97US-0001737.  
 PR  
 XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.  
 PA  
 XX Mizzen L, Wisniewski J;  
 PI  
 XX WPI; 1999-430397/36.  
 XX P-PSDB; Y23904.  
 DR  
 XX New nucleic acid encoding heat shock protein-60 from Streptococcus,  
 PT useful in vaccines, as carriers for other immunogens, as anticancer  
 PT agents and for diagnosis

XX Claim 3; Fig 4A-B; 176pp; English.  
 XX  
 CC The present sequence encodes a heat shock protein, designated Hsp60-2.  
 CC The protein, its fragments, variants and fusion proteins, are  
 CC used to elicit or enhance an immune response against Streptococcus,  
 CC and to elicit a similar response to a target antigen fused to the  
 CC protein. Unlike other immunological carriers, Hsp60 proteins are not  
 CC immunosuppressive so provide an increased response to any conjugated or  
 CC fused antigen. Also, where used for cancer control, they lack the side  
 CC effects associated with endotoxins. They can also be used to detect  
 CC specific antibodies and in treatment or prevention of tumours  
 CC (e.g. sarcoma or cancers of breast, ovary, prostate, lung, pancreas or  
 CC liver). The Hsp60 polynucleotide is used for recombinant production  
 CC of the protein, as a source of primers and probes for detecting  
 CC streptococci in standard hybridization/amplification assays, and  
 CC therapeutically in gene therapy vectors.  
 XX  
 SQ Sequence 1661 BP; 512 A; 308 C; 407 G; 434 T; 0 other;

Query Match 63.6%; Score 14; DB 20; Length 1661;  
 Best Local Similarity 63.6%; Pred. No. 57;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttttgannnnnnnttgata 22  
 |||||  
 DB 1163 TGTGGAGCTCTACTTTGATA 1142

RESULT 7  
 Q99430/C  
 ID Q99430 standard; DNA; 4197 BP.  
 XX AC Q99430;  
 XX  
 DT 22-DEC-1995 (first entry)  
 XX  
 DE B. sphaericus SLP gene.  
 XX  
 KW Surface layer protein; SLP; fusion protein; vaccine; antigen;  
 KW surface expression; epitope; ds.  
 XX  
 OS Bacillus sphaericus.

FH key Location/Qualifiers  
 FT RBS 79..85  
 FT /\*tag= a  
 FT CDS 95..3853  
 FT /\*tag= b  
 FT sig\_peptide 95..184  
 FT /\*tag= c  
 FT mat\_peptide 185..3850  
 FT /\*tag= d  
 XX  
 PN W09519371-A2.  
 XX  
 PD 20-JUL-1995.  
 XX  
 PF 13-JAN-1995; 95WO-EP00147.  
 XX  
 PR 14-JAN-1994; 94GB-0000650.  
 XX  
 PA (SOLV ) SOLVAY SA.

PI Deblaere RY, Desomer J, Dhaese P;  
 XX  
 DR WPI; 1995-263827/34.  
 DR P-PSDB; R80530.  
 XX  
 PT Host cell expressing surface layer protein fusion protein - used for  
 PT host presentation of antigens and vaccine prodn.  
 XX

PS Disclosure; Fig.6; 95pp; English.

XX  
 CC A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG  
 CC P-13855), surface layer protein was used to screen an HindIII-  
 CC generated library to isolate the slp gene. Promoter regions  
 CC of the gene are used in genetic constructs providing surface  
 CC expression of heterologous proteins in P-1 hosts.  
 XX  
 SQ Sequence 4197 BP; 1470 A; 731 C; 763 G; 1233 T; 0 other;

Query Match 63.6%; Score 14; DB 16; Length 4197;  
 Best Local Similarity 63.6%; Pred. No. 63;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttttgannnnnnnttgata 22  
 |||||  
 DB 2836 TGTGGAGCACTGGTTTGATA 2815

RESULT 8  
 X34752  
 ID X34752 standard; DNA; 396 BP.  
 XX  
 AC X34752;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE DNA encoding OMP-1T protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN W09913720-A1.  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 DR P-PSDB; Y06952.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Disclosure; Fig 12A; 55pp; English.

XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in Y06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in Y06959-970. The proteins and genes are used  
 CC to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 396 BP; 147 A; 49 C; 58 G; 142 T; 0 other;

Query Match 59.1%; Score 13; DB 20; Length 396;  
 Best Local Similarity 61.9%; Pred. No. 1.7e+02;  
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 gttggannnnnnnttgata 22  
 |||||  
 DB 193 gttggaatgagttattgata 213

RESULT 9  
 Z16735  
 ID 216735 standard; cDNA; 828 BP.  
 XX  
 AC  
 Z16735;  
 XX  
 DT 12-OCT-1999 (first entry)  
 XX  
 DE Human gene expression product cDNA sequence SEQ ID NO:4205.  
 XX  
 KW Human; gene; gene expression product; diagnosis; therapy; probe;  
 KW detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0938972-A2.  
 XX  
 PD 05-AUG-1999.  
 XX  
 PF 28-JAN-1999; 99WO-US01619.  
 XX  
 PR 03-APR-1998; 98US-0080666.  
 PR 28-JAN-1998; 98US-0072910.  
 PR 24-FEB-1998; 98US-0075954.  
 PR 31-MAR-1998; 98US-0080114.  
 PR 03-APR-1998; 98US-0080515.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
 XX  
 DR WPI; 1999-494092/41.  
 XX  
 PT Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 XX  
 PS Claim 1; Page 1992; 2479pp; English.  
 XX  
 CC The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in 212532 to 217779. Also described is a  
 CC method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in 212532 to 217779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.  
 XX  
 SQ Sequence 828 BP; 189 A; 155 C; 172 G; 227 T; 85 other;

Query Match 59.1%; Score 13; DB 20; Length 828;  
 Best Local Similarity 59.1%; Pred. No. 1.8e+02;  
 Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnnntttgata 22  
 |||||  
 Db 199 tgttgagaattcatttgnta 220  
 |||||  
 RESULT 10  
 X20676/c  
 ID X20676 standard; DNA; 1079 BP.  
 XX  
 AC X20676;  
 XX  
 DT 05-MAY-1999 (first entry)  
 XX  
 DE Polynucleotide sequence from the genome of Treponema pallidum.  
 XX  
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;  
 KW enzyme production; ds.  
 XX  
 OS Treponema pallidum.  
 XX  
 PN W09859034-A2.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 23-JUN-1998; 98WO-US13041.  
 XX  
 PR 24-JUN-1997; 97US-0050667.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Fraser CM;  
 XX  
 DR WPI; 1999-081273/07.  
 XX  
 PT New isolated Treponema pallidum nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of T. pallidum infections, particularly syphilis  
 XX  
 PS Claim 1; Page 842-843; 1150pp; English.  
 XX  
 CC X20500-21243 represent polynucleotide sequences from the genome of  
 CC Treponema pallidum. The sequences can be used for detection,  
 CC diagnosis, characterisation, prevention and therapy for T. pallidum  
 CC infections, particularly syphilis. They can also be used for detecting  
 CC diseases related to Borrelia infections in animals, and for the  
 CC production of biosynthetic products such as enzymes.  
 XX  
 SQ Sequence 1079 BP; 397 A; 280 C; 181 G; 221 T; 0 other;

Query Match 59.1%; Score 13; DB 20; Length 1079;  
 Best Local Similarity 61.9%; Pred. No. 1.9e+02;  
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 gttggannnnnnntttgata 22  
 |||||  
 Db 1022 GTTGGATTAAAGGTGTTGATA 1002  
 |||||  
 RESULT 11  
 V18099  
 ID V18099 standard; DNA; 1482 BP.  
 XX  
 AC V18099;  
 XX  
 DT 04-SEP-1998 (first entry)  
 XX  
 DE Nucleotide sequence of breakpoint region on chromosome 19.  
 XX  
 KW Hydronephrosis gene; HNG gene; USF2 gene; renal disease; renal aplasia;  
 KW vesical-ureteral reflux; pelvi-ureteral junction obstruction;  
 KW multicystic renal dysplasia; renal agenesis; hydronephrosis;

KW Von Mayer-Rokitansky-Kuester disorder; bifid ureter; ss.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..1222  
 FT /\*tag= a  
 FT /\*note= "sequences derived from chromosome 6p21"  
 FT Intron 1..22  
 FT /\*tag= b  
 FT /\*note= "partial intronic sequence of HNG"  
 FT exon 23..170  
 FT /\*tag= c  
 FT /\*number= 9  
 FT /\*note= "exon 9 of HNG"  
 FT misc\_feature 356..533  
 FT /\*tag= d  
 FT /\*note= "Alu repeats"  
 FT misc\_feature 1223..1482  
 FT /\*tag= e  
 FT /\*note= "chromosome 19-derived USF2 sequences"  
 FT exon 1466..1477  
 FT /\*tag= f  
 FT /\*number= 7  
 FT /\*note= "partial sequence of exon 7 of USF2 gene;  
 FT the transcriptional orientation is from the  
 FT centromere to the telomere"  
 FT  
 PN W09815650-A2.  
 XX  
 XX 16-APR-1998.  
 XX  
 PF 09-OCT-1997; 97WO-EP05583.  
 XX  
 PR 09-OCT-1996; 96EP-0202820.  
 XX  
 PA (VLAAS) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.  
 PI Frys JGJ, Groenen PMA, Van De Ven WJM;  
 XX WPI; 1998-240833/21.  
 DR  
 XX Hydronephrosis gene - useful to treat or diagnose renal diseases and  
 PT disorders, e.g. vesical-ureteral reflux, pelvi-ureteral junction  
 PT obstruction, multicystic renal dysplasia or renal agenesis  
 XX  
 PS Disclosure; Fig 9-3; 73pp; English.  
 XX  
 CC This represents the nucleotide sequence of the breakpoint region on  
 CC der(19) which reveals that the transcriptional orientation of the  
 CC hydronephrosis (HNG) gene on the short arm of chromosome 6 is from the  
 CC telomere to the centromere. A translocation partner to this gene on  
 CC chromosome 6 is the chromosome 19 USF2 gene. The HNG gene can be used  
 CC as a starting point to design suitable compounds or techniques for the  
 CC treatment of renal diseases or disorders, or nucleotide probes for  
 CC diagnosing cells involved in renal diseases or disorders. A protein or  
 CC a fragment encoded by HNG gene can be used as a starting point for  
 CC preparing suitable antibodies for diagnosing cells involved in renal  
 CC diseases and disorders. The products and method can be used to treat or  
 CC diagnose renal diseases and disorders selected from vesical-ureteral  
 CC reflux, uni or bilateral pelvi-ureteral junction obstruction, multicystic  
 CC renal dysplasia, renal agenesis, renal aplasia, hydronephrosis,  
 CC Von Mayer-Rokitansky-Kuester disorder and bifid ureter.  
 XX  
 SQ Sequence 1482 BP; 423 A; 310 C; 288 G; 460 T; 1 other;

Query Match 59.1%; Score 13; DB 19; Length 1482;  
 Best Local Similarity 61.9%; Pred. No. 2e+02;  
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 OY 1 tggtagannnnnnntttgat 21  
 |||||||

Db 268 tgttgagtttctctttgat 288  
 RESULT 12  
 T60350/c  
 ID T60350 standard; cDNA to mRNA; 1786 BP.  
 XX  
 AC T60350;  
 XX  
 DT 04-JUN-1997 (first entry)  
 XX  
 DE MAP kinase #2 coding sequence.  
 XX  
 KW Mitogen activated protein kinase; MAP; infectious specific protein;  
 KW plant; jasmone acid; ss.  
 XX  
 OS Synthetic.  
 XX JP09065881-A.  
 PN 11-MAR-1997.  
 XX  
 PF 29-AUG-1995; 95JP-0220935.  
 XX  
 PR 29-AUG-1995; 95JP-0220935.  
 XX  
 PA (NORQ ) NORINSUISANSHO NOGYO SEIBUTSU SHIGEN.  
 XX  
 DR WPI; 1997-220416/20.  
 DR P-PSDB; W15513.  
 XX  
 PT Wound-stress inducible MAP kinase - used to regulate the synthesis  
 PT of jasmone acid  
 XX  
 PS Disclosure; Page 14-16; 21pp; Japanese.  
 XX  
 CC T60349 and T60350 represent the coding sequences for the mitogen  
 CC activated protein (MAP) kinases of the invention. The protein encoded by  
 CC this sequence contains the T-E-Y sequence, in which the threonine and  
 CC tyrosine residues are phosphorylated to activate the protein at residues  
 CC 249-251. The MAP kinase and its gene can be used to regulate the  
 CC synthesis of jasmone acid and the synthesis of a group of infectious  
 CC specific proteins. By introducing the MAP kinase gene into a plant,  
 CC thereby inducing expression of the mRNA for MAP kinase (or its antisense  
 CC RNA), the synthesis of jasmone acid and a group of infectious specific  
 CC proteins can be regulated in the plant.  
 XX  
 SQ Sequence 1786 BP; 503 A; 390 C; 353 G; 540 T; 0 other;

Query Match 59.1%; Score 13; DB 18; Length 1786;  
 Best Local Similarity 61.9%; Pred. No. 2e+02;  
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 OY 1 tggtagannnnnnntttgat 21  
 |||||||  
 Db 116 TGTTCGATATTTTATTGAT 96  
 RESULT 13  
 X07102  
 ID X07102 standard; DNA; 2190 BP.  
 XX  
 AC X07102;  
 XX  
 DT 21-MAY-1999 (first entry)  
 XX  
 DE Staphylococcus aureus mutant p15C31 virulence gene.  
 XX  
 KW Virulence; p7C18; vaccine; antibacterial; antibiotic;  
 KW screening; phosphoribosylformylglycinamide decarboxylase; PurL; ss.  
 XX  
 OS Staphylococcus aureus.



XX WO9901473-A2.  
 XX 14-JAN-1999.  
 XX  
 XX 03-JUL-1998; 98WO-GB01974.  
 XX  
 XX 03-JUL-1997; 97US-0887534.  
 XX  
 XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX  
 XX Holden DW;  
 XX  
 XX WPI; 1999-105999/09.  
 XX P-PSDB; W97694.  
 XX  
 XX Inhibition of virulence genes from *Staphylococcus aureus* - useful  
 XX for, e.g. screening for potential anti-microbial agents  
 XX  
 XX Claim 1; Page 86-89; 203pp; English.  
 XX  
 XX This is the nucleotide sequence of a virulence gene identified in  
 CC *Staphylococcus aureus* mutant P15C31. *S. aureus* genes (see X07088-136)  
 CC associated with virulence were identified by signature-tagged  
 CC mutagenesis in which mutants containing a chromosomal insertion of  
 CC a signature tagged transposon were generated, mutants with  
 CC attenuated virulence were identified in a mouse model of bacteraemia,  
 CC and the nucleotide sequences of the regions flanking the transposon  
 CC insertion sites of these mutants were determined. Database sequence  
 CC comparisons were performed to identify the virulence genes and to  
 CC determine the possible function of their protein products (see also  
 CC W97680-724). The P15C31 virulence gene product (see W97694) was  
 CC identified as phosphoribosylformylglycinamide decarboxylase PurL,  
 CC an enzyme involved in purine biosynthesis. A claimed method of  
 CC identifying an antibacterial agent involves assaying potential agents  
 CC for the ability to interfere with the expression of *S. aureus*  
 CC virulence gene products. Also new is a *S. aureus* organism containing  
 CC a functional mutation in one of the virulence genes, and its use in  
 CC vaccine compositions.  
 XX  
 XX Sequence 2190 BP; 753 A; 309 C; 461 G; 666 T; 1 other;  
 SQ

Query Match 59.1%; Score 13; DB 20; Length 2190;  
 Best Local Similarity 61.9%; Pred. No. 2.1e+02;  
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnnntttgat 21  
 ||||| |||||  
 Db 1650 tgttgaatgtaggtttgat 1670

RESULT 14  
 T67197  
 ID T67197 standard; cDNA; 2508 BP.  
 XX  
 XX AC T67197;  
 XX  
 XX 03-SEP-1997 (first entry)  
 XX  
 XX Zebrafish retinoid X receptor RXR epsilon cDNA.  
 DE  
 DE Zebrafish retinoid X receptor RXR epsilon cDNA.  
 KW Retinoid X receptor; RXR epsilon; zebrafish; ds.  
 XX  
 XX Brachydanio rerio.  
 OS  
 OS Location/Qualifiers  
 FH Key 388..1704  
 FT CDS /tag= a  
 FT  
 FT CA2177642-A.  
 XX  
 XX  
 XX 06-DEC-1996.  
 PD

XX 29-MAY-1996; 96CA-2177642.  
 XX  
 XX 05-JUN-1995; 95US-0462182.  
 XX  
 XX (TOOH ) UNIV QUEENS KINGSTON.  
 XX  
 XX Jones B, Ohno C, Petkovich M;  
 XX  
 XX WPI; 1997-298660/28.  
 XX P-PSDB; W18033.  
 XX  
 XX Isolated retinoid X receptor protein - having amino acid insert in  
 XX ligand binding domain, useful to identify specific target genes  
 XX implicated in retinoid responses important in disease states  
 XX  
 XX Claim 8; Page 29-30; 42pp; English.  
 XX  
 XX cDNA clones (T67196 and T67197) respectively code for novel  
 CC retinoid X receptors RXR delta (W18032) and RXR epsilon (W18033)  
 CC that exhibit a high degree of amino acid conservation with other  
 CC vertebrate RXRs but which represent unique subtypes defined by an  
 CC additional 14-amino acid segment in their ligand binding domains.  
 CC The RXR epsilon clone was isolated by screening adult and post-  
 CC somitogenesis zebrafish cDNA libraries with a probe corresponding  
 CC to the *Drosophila* melanogaster FTZ-F1alpha DNA binding domain.  
 CC The isolated clones can be used to produce RXR delta and epsilon  
 CC polypeptides useful for modifying retinoid activity and identifying  
 CC specific target genes implicated in retinoid responses important in  
 CC disease states.  
 XX  
 XX Sequence 2508 BP; 600 A; 652 C; 646 G; 610 T; 0 other;  
 SQ

Query Match 59.1%; Score 13; DB 18; Length 2508;  
 Best Local Similarity 61.9%; Pred. No. 2.1e+02;  
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnnntttgat 21  
 ||||| |||||  
 Db 2424 tgttgacatgagttttgat 2444

RESULT 15  
 Z94941/c  
 ID Z94941 standard; cDNA; 3519 BP.  
 XX  
 XX AC Z94941;  
 XX  
 XX 01-AUG-2000 (first entry)  
 XX  
 XX Human carbohydrate-associated protein CRBAP-1 cDNA.  
 DE  
 DE CRBAP-1; carbohydrate-associated protein 1; human;  
 KW autoimmune disorder; inflammation; gastrointestinal disorder;  
 KW infection; reproductive disorder; neurological disorder;  
 KW eye disorder; cell proliferation; cancer; diagnosis; gene therapy;  
 KW ss.  
 XX  
 XX Homo sapiens.  
 OS  
 OS Key Location/Qualifiers  
 FH CDS 42..560  
 FT /tag= a  
 FT variation replace(428,C)  
 FT /tag= b  
 FT /frequency= "0.10"  
 FT /note= "a polynucleotide containing C at position  
 FT 428 is specifically claimed in Claimed 21"  
 XX  
 XX WO200018922-A2.  
 XX  
 XX 06-APR-2000.  
 PD

```

XX 29-SEP-1999; 99WO-US22685.
XX 01-OCT-1998; 98US-0164785.
XX 06-OCT-1998; 98US-0167179.
XX 13-NOV-1998; 98US-0191838.
XX 03-DEC-1998; 98US-0205856.
XX 03-DEC-1998; 98US-2223333.
XX (INCY-) INCYTE PHARM INC.
XX Au-Young J, Lal P, Bandman O, Reddy R, Baughn MR, Yue H;
XX Hillman JL;
XX WPI: 2000-317516/27.
XX P-PSDB: Y79505.
XX Novel carbohydrate-associated proteins used for the prevention and
XX treatment of autoimmune/inflammatory disorders of e.g. the
XX gastrointestinal and reproductive systems -
XX Claim 9: Page 90-91; 104pp; English.
XX The present sequence is that of cDNA coding for a novel human
XX carbohydrate-associated protein, termed CRBAP-1 (see Y79505).
XX The cDNA (Incyte clone 714029) was initially identified in
XX prostate tumour cDNA library PROSTUT01. The sequence contains a
XX single nucleotide polymorphism with C replacing T at position
XX 428. C was found at position 428 in 10% of the clones, and
XX T in 90% of clones. CRBAP-1 resembles vertebrate galactoside-binding
XX lectins, showing 28% identity to human galectin-8. CRBAP-1 expression
XX was identified in libraries associated with cancer and cell
XX proliferation, inflammation and immune response, reproductive tissues
XX and nervous tissues. The invention provides CRBAP-1 to -7
XX polynucleotides (see Z94941-48) and polypeptides (see Y79505-11),
XX as well as expression vectors, host cells, antibodies, agonists
XX and antagonists. These are used in the diagnosis, treatment or
XX prevention of disorders associated with CRBAP expression,
XX especially autoimmune or inflammatory disorders, gastrointestinal
XX disorders, infectious disorders, reproductive disorders,
XX neurological disorders, eye disorders and cell proliferative
XX disorders, including cancer. CRBAP polynucleotides are useful
XX sources of probes and primers which can be used to detect CRBAP in
XX a sample from a patient. They may also be administered as part of
XX a gene therapy regime.
XX Sequence 3519 BP; 1067 A; 627 C; 730 G; 1095 T; 0 Other;
SQ
Query Match 59.1%; Score 13; DB 21; Length 3519;
Best Local Similarity 61.9%; Pred. No. 2.2e+02;
Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 tgttggannnnnnnttgat 21
Db 1909 TGTGGAGTCACCTGTTTCAT 1889

```

Search completed: March 27, 2001, 08:23:01  
Job time: 5303 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:22 ; Search time 132.3 Seconds  
(without alignments)  
26.799 Million cell updates/sec

Title: US-09-380-826A-6  
Perfect score: 22  
Sequence: 1 ttttgannnnnnnttgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
C 1	14	63.6	3666	2	US-08-682-517-13 Sequence 13, Appl
C 2	14	63.6	3666	2	US-08-682-517-14 Sequence 14, Appl
C 3	14	63.6	4197	2	US-08-682-517-7 Sequence 7, Appl
C 4	14	63.6	4197	2	US-08-682-517-8 Sequence 8, Appl
C 5	13	59.1	5253	3	US-08-714-918-19 Sequence 19, Appl
C 6	12.4	56.4	1760	1	US-08-413-118-118 Sequence 118, App
C 7	12.4	56.4	1760	3	US-08-473-446-118 Sequence 118, App
C 8	12.4	56.4	1800	1	US-08-752-238-2 Sequence 2, Appl
C 9	12.4	56.4	1800	3	US-09-085-603B-2 Sequence 2, Appl
C 10	12.4	56.4	1800	3	US-09-031-897-6 Sequence 6, Appl
C 11	12.4	56.4	2730	1	US-08-339-129-1 Sequence 1, Appl
C 12	12.4	56.4	2761	1	US-08-752-238-1 Sequence 1, Appl
C 13	12.4	56.4	2761	3	US-09-085-603B-1 Sequence 1, Appl
C 14	12.4	56.4	2761	3	US-09-031-897-5 Sequence 5, Appl
C 15	12.4	56.4	3438	3	US-08-613-009A-1 Sequence 1, Appl
C 16	12	54.5	1130	1	US-07-864-004B-1 Sequence 1, Appl
C 17	12	54.5	1130	1	US-08-251-937A-1 Sequence 1, Appl
C 18	12	54.5	1130	1	US-08-212-133A-5 Sequence 5, Appl
C 19	12	54.5	1130	1	US-08-474-503-3 Sequence 3, Appl
C 20	12	54.5	1130	2	US-08-670-707A-3 Sequence 3, Appl
C 21	12	54.5	1130	4	PCT-US93-03275-1 Sequence 1, Appl
C 22	12	54.5	1130	4	PCT-US94-13200-3 Sequence 3, Appl
C 23	12	54.5	1623	1	US-08-121-202-3 Sequence 3, Appl
C 24	12	54.5	2026	3	US-08-755-587-26 Sequence 26, Appl
C 25	12	54.5	2625	1	US-08-468-036-2 Sequence 2, Appl
C 26	12	54.5	2625	2	US-08-376-843-2 Sequence 2, Appl
C 27	12	54.5	3284	3	US-09-136-652-1 Sequence 1, Appl
C 28	12	54.5	4334	2	US-08-670-707A-38 Sequence 38, Appl

C 29 12 54.5 4931 3 US-09-058-489-20 Sequence 20, Appl  
C 30 12 54.5 6057 3 US-08-362-525-1 Sequence 1, Appl  
C 31 12 54.5 6402 2 US-08-670-707A-36 Sequence 36, Appl  
C 32 12 54.5 6476 3 US-09-058-489-21 Sequence 21, Appl  
C 33 12 54.5 6727 3 US-08-629-643A-5 Sequence 5, Appl  
C 34 12 54.5 6727 3 US-09-280-799-1 Sequence 1, Appl  
C 35 12 54.5 13011 2 US-08-791-849A-14 Sequence 14, Appl  
C 36 12 54.5 40352 3 US-08-846-111D-15 Sequence 15, Appl  
C 37 11.4 51.8 21 2 US-08-639-501-23 Sequence 23, Appl  
C 38 11.4 51.8 21 3 US-09-044-946-23 Sequence 23, Appl  
C 39 11.4 51.8 21 3 US-09-044-908-23 Sequence 24, Appl  
C 40 11.4 51.8 27 1 US-08-419-009-24 Sequence 24, Appl  
C 41 11.4 51.8 293 3 US-08-866-240-13 Sequence 13, Appl  
C 42 11.4 51.8 336 3 US-09-184-658-11 Sequence 11, Appl  
C 43 11.4 51.8 485 1 US-08-419-009-19 Sequence 19, Appl  
C 44 11.4 51.8 485 1 US-08-419-009-20 Sequence 20, Appl  
C 45 11.4 51.8 485 1 US-08-419-009-21 Sequence 21, Appl

ALIGNMENTS

RESULT 1  
US-08-682-517-13/c.  
; Sequence 13, Application US/08682517  
; Patent No. 5874267  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Expression of surface layer proteins  
; NUMBER OF SEQUENCES: 25  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/682,517  
; FILING DATE:  
; CLASSIFICATION:  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3666 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-682-517-13

Query Match 63.6%; Score 14; DB 2; Length 3666;  
Best Local Similarity 63.6%; Pred. No. 25;  
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttttgannnnnnnttgata 22  
Db 2652 TGTGGAGCAACTGGTTTGATA 2631

RESULT 2  
US-08-682-517-14/c  
; Sequence 14, Application US/08682517  
; Patent No. 5874267  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Expression of surface layer proteins  
; NUMBER OF SEQUENCES: 25  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/682,517

;; FILING DATE:  
;; CLASSIFICATION:  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3666 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..3666  
US-08-682-517-14

Query Match 63.6%; Score 14; DB 2; Length 3666;  
Best Local Similarity 63.6%; Pred. No. 25;  
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttttgannnnnnnnnttgata 22  
|||||||  
DB 2652 TGTGGAGCAACTGGTTTGATA 2631

## RESULT 3

US-08-682-517-7/c  
; Sequence 7, Application US/08682517  
; Patent No. 5874267  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Expression of surface layer proteins  
; NUMBER OF SEQUENCES: 25  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/682,517  
; FILING DATE:  
; CLASSIFICATION:  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4197 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-682-517-7

Query Match 63.6%; Score 14; DB 2; Length 4197;  
Best Local Similarity 63.6%; Pred. No. 25;  
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttttgannnnnnnnnttgata 22  
|||||||  
DB 2836 TGTGGAGCAACTGGTTTGATA 2815

## RESULT 4

US-08-682-517-8/c  
; Sequence 8, Application US/08682517  
; Patent No. 5874267  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Expression of surface layer proteins  
; NUMBER OF SEQUENCES: 25  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/682,517  
;; FILING DATE:  
;; CLASSIFICATION:  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4197 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORIGINAL SOURCE:  
;; ORGANISM: Bacillus sphaericus  
;; INDIVIDUAL ISOLATE: P-1  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 95..3850  
;; FEATURE:  
;; NAME/KEY: mat\_peptide  
;; LOCATION: 185..3850  
;; FEATURE:  
;; NAME/KEY: sig\_peptide  
;; LOCATION: 95..184  
US-08-682-517-8

Query Match 63.6%; Score 14; DB 2; Length 4197;  
Best Local Similarity 63.6%; Pred. No. 25;  
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttttgannnnnnnnnttgata 22  
|||||||  
DB 2836 TGTGGAGCAACTGGTTTGATA 2815

## RESULT 5

US-08-714-918-19  
; Sequence 19, Application US/08714918  
; Patent No. 6037123  
; GENERAL INFORMATION:  
; APPLICANT: Benton, Bret  
; APPLICANT: Lee, Ving  
; APPLICANT: Malouin, Francois  
; APPLICANT: Martin, Patrick K.  
; APPLICANT: Schmid, Molly B.  
; APPLICANT: Sun, Dongxu  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL  
; TARGET GENES  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C.-DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/714,918  
; FILING DATE: September 13, 1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/009,102  
; FILING DATE: December 22, 1995  
; APPLICATION NUMBER: 60/003,798  
; FILING DATE: September 15, 1995  
; ATTORNEY/AGENT INFORMATION:

Query Match 56.4%; Score 12.4; DB 1; Length 1760;  
Best Local Similarity 59.1%; Pred. No. 1.7e+02;

NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: US  
ZIP: 44114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,238  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Golrick, Mary E.  
REGISTRATION NUMBER: 34,829  
REFERENCE/DOCKET NUMBER: 24671/00103  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (216) 622-8458  
TELEFAX: (216) 241-0816  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1800 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1800  
US-08-752-238-2

Query Match 56.4%; Score 12.4; DB 1; Length 1800;  
Best Local Similarity 59.1%; Pred. No. 1.7e+02;  
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ttttggannnnnnnttgata 22  
| | | | |  
Db 139 TATTGGAATATAAATTTTGATA 118

RESULT 9  
US-09-085-603B-2/c  
Sequence 2, Application US/09085603B  
Patent No. 6001608  
GENERAL INFORMATION:  
APPLICANT: Lambowitz Dr., Alan M  
APPLICANT: Mohr Dr., Georg  
APPLICANT: Saldanha Dr., Roland  
APPLICANT: Matsuura Dr., Manabu  
APPLICANT: Yang Dr., Jiam  
APPLICANT: Zimmerly Dr., Steven  
APPLICANT: Guo Dr., Huatao  
APPLICANT: Beall Dr., Clifford J.  
TITLE OF INVENTION: Methods of making an Rnp Particle  
TITLE OF INVENTION: Having Nucleotide Integrase Activity  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD LLP  
STREET: 800 Superior Avenue  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: US  
ZIP: 44114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/085,603B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Docherty, Pamela A.  
REGISTRATION NUMBER: 40,591  
REFERENCE/DOCKET NUMBER: 24671/04000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (216) 622-8416  
TELEFAX: (216) 241-0816  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1800 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1800  
US-09-085-603B-2  
Query Match 56.4%; Score 12.4; DB 3; Length 1800;  
Best Local Similarity 59.1%; Pred. No. 1.7e+02;  
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ttttggannnnnnnttgata 22  
| | | | |  
Db 139 TATTGGAATATAAATTTTGATA 118

RESULT 10  
US-09-031-897-6/c  
Sequence 6, Application US/09031897  
Patent No. 6027895  
GENERAL INFORMATION:  
APPLICANT: Lambowitz, Alan  
APPLICANT: Mohr, Georg  
APPLICANT: Zimmerly, Steven  
APPLICANT: Guo, Huatao  
TITLE OF INVENTION: Methods Cleaving DNA with Nucleotide  
TITLE OF INVENTION: Integrases  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calfee, Halter & Griswold  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: US  
ZIP: 44114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/031,897  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Docherty, Pamela A.  
REGISTRATION NUMBER: 40,591  
REFERENCE/DOCKET NUMBER: 24671/00105  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (216) 622-8416  
TELEFAX: (216) 241 0816  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1800 base pairs  
TYPE: nucleic acid

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1800
US-09-031-897-6

Query Match          56.4%; Score 12.4; DB 3; Length 1800;
Best Local Similarity 59.1%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
| ||||| |||||
Db 139 TATTGGAATATAAATTGTGATA 118

RESULT 11
US-08-339-129-1/c
; Sequence 1, Application US/08339129
; Patent No. 5750399
; GENERAL INFORMATION:
; APPLICANT: Dixon, Richard A.
; APPLICANT: Paiva, Nancy L.
; APPLICANT: Commen, Abraham
; TITLE OF INVENTION: Isoflavone Reductase Promoter
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richards, Medlock & Andrews
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: Texas
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,129
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Hansen, Eugenia S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: NOBF B35969
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-939-4500
; TELEFAX: 214-939-4600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-339-129-1

Query Match          56.4%; Score 12.4; DB 1; Length 2730;
Best Local Similarity 59.1%; Pred. No. 1.8e+02;
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
| ||||| |||||
Db 950 TGTGGATTTCCTGCTTTATA 929

RESULT 12
US-08-752-238-1/c
; Sequence 1, Application US/08752238
; Patent No. 5804418
; GENERAL INFORMATION:
; APPLICANT: Lambowitz Dr., Alan M
; APPLICANT: Mohr Dr., Georg
; APPLICANT: Saldanha Dr., Roland
; APPLICANT: Matsuura Dr., Manabu
; APPLICANT: Yang Dr., Jiam
; APPLICANT: Zimmerly Dr., Steven
; APPLICANT: Guo Dr., Huatao
; APPLICANT: Beall Dr., Clifford J.
; TITLE OF INVENTION: Methods of Making an Rnp Particle
; TITLE OF INVENTION: Having Nucleotide Integrase Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CALFEE, HALTER & GRISWOLD LLP
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: US

; Patent No. 5804418
; GENERAL INFORMATION:
; APPLICANT: Lambowitz Dr., Alan M
; APPLICANT: Mohr Dr., Georg
; APPLICANT: Saldanha Dr., Roland
; APPLICANT: Matsuura Dr., Manabu
; APPLICANT: Yang Dr., Jiam
; APPLICANT: Zimmerly Dr., Steven
; APPLICANT: Guo Dr., Huatao
; APPLICANT: Beall Dr., Clifford J.
; TITLE OF INVENTION: Methods of Making an Rnp Particle
; TITLE OF INVENTION: Having Nucleotide Integrase Activity
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CALFEE, HALTER & GRISWOLD
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: US

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,238
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goltick, Mary E.
; REGISTRATION NUMBER: 34,829
; REFERENCE/DOCKET NUMBER: 24671/00103
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8458
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2761 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-752-238-1

Query Match          56.4%; Score 12.4; DB 1; Length 2761;
Best Local Similarity 59.1%; Pred. No. 1.8e+02;
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 tgttgannnnnnnttgata 22
| ||||| |||||
Db 890 TATTGGAATATAAATTGTGATA 869

RESULT 13
US-09-085-603B-1/c
; Sequence 1, Application US/09085603B
; Patent No. 6001608
; GENERAL INFORMATION:
; APPLICANT: Lambowitz Dr., Alan M
; APPLICANT: Mohr Dr., Georg
; APPLICANT: Saldanha Dr., Roland
; APPLICANT: Matsuura Dr., Manabu
; APPLICANT: Yang Dr., Jiam
; APPLICANT: Zimmerly Dr., Steven
; APPLICANT: Guo Dr., Huatao
; APPLICANT: Beall Dr., Clifford J.
; TITLE OF INVENTION: Methods of Making an Rnp Particle
; TITLE OF INVENTION: Having Nucleotide Integrase Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CALFEE, HALTER & GRISWOLD LLP
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: US
```

```
; LENGTH: 2761 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-031-897-5

Query Match          56.4%; Score 12.4; DB 3; Length 2761;
Best Local Similarity 59.1%; Pred. No. 1.8e+02;
Matches 13; Conservative 0; Mismatches 9; Indels 0

Qy      1 tggttgagannnnnnnttggata 22
        | |||||                |||||
Db      890 TATTGGAATATAAAATTTTGATA 869

RESULT 15
US-08-613-009A-1/c
; Sequence 1, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
```

```

:
:
: APPLICANT: Du, Run-Pan
: APPLICANT: Yang, Yan-Ping
: APPLICANT: Klein, Michel H
: TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McGurney
:

```

```

; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: .08-Mar-1996
; APPLICATION NUMBER: US/08/613,009A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3438 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-613-009A-1

Query Match 56.4%; Score 12.4; DB 3; Length 3438
Best Local Similarity 59.1%; Pred. No. 1.9e+02;
Matches 13; Conservative 0; Mismatches 9; Indels

Oy 1 tgttgannnnnnnttgata 22
|||||
Db 202 TGTGTGCTTTTGCACTTGATA 181

```



Search completed: March 27, 2001, 08:19:23  
Job time: 5150 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:23 ; Search time 2517.78 Seconds  
(without alignments)  
61.230 Million cell updates/sec

Title: US-09-380-826A-6  
Perfect score: 22  
Sequence: 1 tgttgannnnnnntttgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estrol:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est58:\*  
105: gb\_est59:\*  
106: gb\_est60:\*  
107: gb\_est61:\*  
108: gb\_est62:\*  
109: gb\_est63:\*  
110: gb\_est64:\*  
111: gb\_est65:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estpl6:\*  
 118: em\_estpl7:\*  
 119: em\_estpl8:\*  
 120: em\_estpl9:\*  
 121: em\_estpl10:\*  
 122: em\_estpl11:\*  
 123: em\_estpl12:\*  
 124: em\_estpl13:\*  
 125: em\_estpl14:\*  
 126: em\_estpl15:\*  
 127: em\_estpl16:\*  
 128: em\_estpl17:\*  
 129: em\_estpl18:\*  
 130: em\_estpl19:\*  
 131: em\_estpl20:\*  
 132: em\_estpl21:\*  
 133: em\_estpl22:\*  
 134: em\_estpl23:\*  
 135: em\_estpl24:\*  
 136: em\_estpl25:\*  
 137: em\_estpl26:\*  
 138: em\_estpl27:\*  
 139: em\_estpl28:\*  
 140: em\_estpl29:\*  
 141: em\_estpl30:\*  
 142: em\_estpl31:\*  
 143: em\_estpl32:\*  
 144: em\_estpl33:\*  
 145: em\_estpl34:\*  
 146: em\_estpl35:\*  
 147: em\_estpl36:\*  
 148: em\_estpl37:\*  
 149: em\_estpl38:\*  
 150: em\_estpl39:\*  
 151: em\_estpl40:\*  
 152: em\_estpl41:\*  
 153: em\_estpl42:\*  
 154: em\_estpl43:\*  
 155: em\_estpl44:\*  
 156: em\_estpl45:\*  
 157: em\_estpl46:\*  
 158: em\_estpl47:\*  
 159: em\_estpl48:\*  
 160: em\_estpl49:\*  
 161: em\_estpl50:\*  
 162: em\_estpl51:\*  
 163: em\_estpl52:\*  
 164: em\_estpl53:\*  
 165: em\_estpl54:\*  
 166: em\_estpl55:\*  
 167: em\_estpl56:\*  
 168: em\_estpl57:\*  
 169: em\_estpl58:\*  
 170: em\_estpl59:\*  
 171: em\_estpl60:\*  
 172: em\_estpl61:\*  
 173: em\_estpl62:\*  
 174: em\_estpl63:\*  
 175: em\_estpl64:\*  
 176: em\_estpl65:\*  
 177: em\_estpl66:\*  
 178: em\_estpl67:\*  
 179: em\_estpl68:\*  
 180: em\_estpl69:\*  
 181: em\_estpl70:\*  
 182: em\_estpl71:\*  
 183: em\_estpl72:\*  
 184: em\_estpl73:\*  
 185: em\_estpl74:\*  
 186: em\_estpl75:\*  
 187: em\_estpl76:\*  
 188: em\_estpl77:\*  
 189: em\_estpl78:\*

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	14	63.6	184	128	BB357781
C 2	14	63.6	210	134	BE059640
C 3	14	63.6	228	138	BE956026
C 4	14	63.6	235	102	BB170254
C 5	14	63.6	250	16	AI122094
C 6	14	63.6	261	96	AW879900
C 7	14	63.6	265	126	BB287901
C 8	14	63.6	298	8	AA515145
C 9	14	63.6	332	105	BE202756
C 10	14	63.6	336	166	AZ037213
C 11	14	63.6	359	37	AV546877
C 12	14	63.6	364	87	AW226052
C 13	14	63.6	371	151	AQ311167
C 14	14	63.6	376	149	AQ102282
C 15	14	63.6	376	163	AQ844476
C 16	14	63.6	378	88	AW240486
C 17	14	63.6	389	8	AA488558
C 18	14	63.6	399	144	R89220
C 19	14	63.6	410	93	AW624588
C 20	14	63.6	410	149	AQ150667
C 21	14	63.6	416	9	AA555292
C 22	14	63.6	416	159	AQ549780
C 23	14	63.6	427	23	AI662442
C 24	14	63.6	447	169	AZ273408
C 25	14	63.6	459	136	BE854379
C 26	14	63.6	467	28	AL370583
C 27	14	63.6	471	22	AI591928
C 28	14	63.6	476	164	AQ914606
C 29	14	63.6	481	146	T75994
C 30	14	63.6	482	174	B29252
C 31	14	63.6	489	105	BE196823
C 32	14	63.6	490	108	BE448206
C 33	14	63.6	493	10	AA637362
C 34	14	63.6	496	159	AQ589820
C 35	14	63.6	497	107	BE353412
C 36	14	63.6	498	138	BF003327
C 37	14	63.6	521	93	AW649760
C 38	14	63.6	531	90	AW413577
C 39	14	63.6	531	93	AW648609
C 40	14	63.6	531	160	AQ669223
C 41	14	63.6	536	97	AW945076
C 42	14	63.6	539	27	AI997301
C 43	14	63.6	540	90	AW398351
C 44	14	63.6	544	38	AW678686
C 45	14	63.6	544	169	AZ267308

## ALIGNMENTS

RESULT 1  
 BB357781/C

LOCUS

DEFINITION

musculus CDNA clone C030024G02 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

184 bp

mrna

EST

12-JUL-2000

adult male corpus striatum Mus

house mouse.



## ORIGIN

Query Match 63.6%; Score 14; DB 134; Length 210;  
 Best Local Similarity 63.6%; Pred. No. 6.7e+02;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgagannnnnnnttgata 22  
 |||||  
 Db 43 TGTGGATGTCACACGTTTGATA 64

RESULT 3  
BE956026/c

LOCUS BE956026 228 bp mRNA EST 04-OCT-2000  
 DEFINITION UI-M-BH4-bav-a-11-0-UI.s1 NIH\_BMAP\_M\_S5 Mus musculus cDNA clone  
 BE956026  
 ACCESSION BE956026.1 GI:10600157  
 VERSION EST.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 228)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery.  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643 USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mestr@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 olfactory bulbs tissue cDNA Library Preparation: M.B. Soares Lab  
 Clone distribution: Researchers may obtain BMAP cDNA clones from  
 RESEARCH GENETICS. It should be noted that Bento Soares is  
 generating a small number of additional specialized non-redundant  
 arrays of BMAP cDNAs whose availability will be considered under  
 appropriate and limited collaborative arrangements  
 Seq primer: M13 Forward  
 POLYA-Tes.

## FEATURES

## source

Location/Qualifiers  
 1..228  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH4-bav-a-11-0-UI"  
 /clone\_lib="NIH\_BMAP\_M\_S5"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker: Site\_1: Not I; Site\_2: Eco RI; The  
 NIH\_BMAP\_M\_S5 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged  
 normalized libraries from ten regions of the mouse brain  
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,  
 cortex, amygdala, basal ganglia, pineal gland, striatum,  
 hippocampus) after a series of subtractions to reduce the  
 representation of cDNAs from which ESTs had already been  
 generated. For a detailed description of the library from  
 which this clone was derived, please visit our web site  
 at brainest.eng.uiowa.edu.  
 TAG\_LIB=NIH\_BMAP\_M\_S5  
 TAG\_TISSUE=olfactory-bulbs

## TAG\_SEQ=CAAGG"

BASE COUNT 64 a 47 c 42 g 75 t  
 ORIGIN

Query Match 63.6%; Score 14; DB 138; Length 228;  
 Best Local Similarity 63.6%; Pred. No. 6.8e+02;  
 Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgagannnnnnnttgata 22  
 |||||  
 Db 144 TGTGGAATCTGTTTGATA 123

RESULT 4  
BB170254/c

LOCUS BB170254 235 bp mRNA EST 29-JUN-2000  
 DEFINITION musculus cDNA clone A230017D13 3', mRNA sequence.  
 BB170254  
 ACCESSION BB170254.1 GI:8829337  
 VERSION EST.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 235)  
 AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
 P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
 Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,  
 Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,  
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
 Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
 Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,  
 T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,  
 Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
 M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al.)  
 UNPUBLISHED (2000)  
 CONTACT: Yoshihide Hayashizaki  
 Genome Exploration Research Group, Life Science Tsukuba Center,  
 The Institute of Physical and Chemical Research (RIKEN), Genomic  
 Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: +81-298-36-9013  
 Fax: +81-298-36-9098  
 Email: genome-res@rtrc.riken.go.jp,  
 URL: http://genome.rtrc.riken.go.jp/  
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Thermotabilization and thermoactivation of thermostable enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh,M., Kitsuina,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
 Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (<http://genome.rtrc.riken.go.jp>) for  
 further details.  
 Location/Qualifiers  
 1..235  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="A230017D13"  
 /clone\_lib="RIKEN full-length enriched, adult male  
 hypothalamus"

FEATURES  
source

/sex="male"  
/tissue\_type="hypothalamus"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/notes="Site 1: Salt; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',  
GAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCTCGATTAAATAATCCGCCGCCGCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 84 a 47 c 31 g 73 t  
ORIGIN

Query Match 63.6%; Score 14; DB 102; Length 235;  
Best Local Similarity 63.6%; Pred. No. 6.9e+02;  
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgagannnnnnnnttgata 22  
||||||| |||||||  
Db 66 TGTGGACTTAGTCTCTTGATA 45

RESULT 5  
LOCUS A1122094  
DEFINITION uc46f10\_r1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone IMAGE:1401067 5' similar to SW:1335\_MTCGE P47577 HYPOTHETICAL GTP-BINDING PROTEIN MG335. ; mRNA sequence.  
ACCESSION A1122094  
VERSION A1122094.1 GI:3522418  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 250)  
AUTHORS Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R. and Waterston R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:912783  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 225.  
Location/Qualifiers

FEATURES  
source  
1. .250  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1401067"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"

/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 48 a 79 c 75 g 48 t  
ORIGIN

Query Match 63.6%; Score 14; DB 16; Length 250;  
Best Local Similarity 63.6%; Pred. No. 6.9e+02;  
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgagannnnnnnnttgata 22  
||||||| |||||||  
Db 211 TGTGGAAAATCCTCTTTGATA 232

RESULT 6  
LOCUS AW879900/c  
DEFINITION QV3-OT0029-290300-134-b02 OT0029 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW879900  
VERSION AW879900.1 GI:8041910  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 261)  
AUTHORS Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R., Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F., Goldman G.H., Carvalho A.F., Matsukuma A., Baia G.S., Simpson D.H., Brunstein A., de Oliveira P.S., Bucher P., Jorgensen C.V., O'Hare M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and Simpson A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl1=6t2-QV3-OT0029-290>)  
300-134-b02&t3=2000-03-29&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 261.  
Location/Qualifiers

FEATURES  
source  
1. .261  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="OT0029"  
/dev\_stage="Adult"  
/note="Organ: ovary; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

•



```

BASE COUNT      76 a      63 c      91 g      68 t
ORIGIN

Query Match      63.6%; Score 14; DB 8; Length 298;
Best Local Similarity 63.6%; Pred. No. 7.2e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22
|||||
DB 5 TGTGGAGATGACATTGATA 26

RESULT 9
BE202756      332 bp      mRNA      EST      07-SEP-2000
LOCUS      EST402778 KV1 Medicago truncatula cDNA clone pkV1-3023, mRNA
DEFINITION      sequence.
ACCESSION      BE202756
VERSION      BE202756
KEYWORDS      EST.
SOURCE      barrel medic.
ORGANISM      Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE      1 (bases 1 to 332)
AUTHORS      Vandenbosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
Fraser,C.M.
TITLE      ESTs from roots of Medicago truncatula 24 hours after inoculation
JOURNAL      with Sinorhizobium meliloti
COMMENT      Unpublished (1999)
Contact: Vandenbosch K
Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel: 409 845 7707
Fax: 409 845 2891
Email: kate@mail.bio.tamu.edu
Texas A&M University:T262522e
TIGR sequence name:MTIAE96TK
More information is available at:
http://chrysie.tamu.edu/medicago
Seq primer: SKnod (CTA gAA CTA gTg gAT CC).
FEATURES
    Location/Qualifiers
    1..332
        /organism="Medicago truncatula"
        /cultivar="genotype A17"
        /db_xref="taxon:3880"
        /clone="pkV1-3023"
        /clone_lib="KV1"
        /tissue_type="Seedling roots"
        /dev_stage="24 hours post-inoculation with Sinorhizobium
        meliloti"
        /lab_host="E. coli strain XL0LR"
        /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
        XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
        was directionally ligated into the Unizap XR vector from
        Stratagene and packaged using Gigapack III Gold packaging
        extracts. Plasmids containing cDNA inserts were excised
        from the recombinant lambda-zap phage using Ex-assist
        helper phage and propagated in XL0LR cells."
BASE COUNT      95 a      50 c      96 g      91 t
ORIGIN

Query Match      63.6%; Score 14; DB 105; Length 332;
Best Local Similarity 63.6%; Pred. No. 7.4e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22
|||||

BASE COUNT      97 a      37 c      45 g      157 t
ORIGIN

Query Match      63.6%; Score 14; DB 166; Length 336;
Best Local Similarity 63.6%; Pred. No. 7.4e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22
|||||
DB 268 TGTGGAAATTTTCCTTTGATA 289

RESULT 11
AV546877
LOCUS      AV546877      359 bp      mRNA      EST      07-SEP-2000
DEFINITION      AV546877 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
CDNA clone R2L21a12F 3', mRNA sequence.
ACCESSION      AV546877

```

```

DB 157 TGTTCGATTGATGGCTTTGATA 178

RESULT 10
AZ037213
LOCUS      AZ037213      336 bp      DNA      GSS      01-MAR-2000
DEFINITION      RPCI-23-364L15.TV RPCI-23 Mus musculus genomic clone RPCI-23-364L15
DNA sequence.
ACCESSION      AZ037213
VERSION      AZ037213.1 GI:7125453
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 336)
AUTHORS      Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Aklnret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and and Fraser,C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL      Unpublished (1999)
COMMENT      Other_GSSs: RPCI-23-364L15.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 364 row: L column: 15
Seq primer: T7
Class: BAC ends.
FEATURES
    Location/Qualifiers
    1..336
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="RPCI-23-364L15"
        /clone_lib="RPCI-23"
        /sex="Female"
        /lab_host="DH10B"
        /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
        EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
        brain genomic DNA was isolated and partially digested
        with a combination of EcoRI and EcoRI Methylase. Size
        selected DNA was cloned into the pBACe3.6 vector at the
        EcoRI sites. The ligation products were transformed into
        DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      97 a      37 c      45 g      157 t
ORIGIN

Query Match      63.6%; Score 14; DB 166; Length 336;
Best Local Similarity 63.6%; Pred. No. 7.4e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22
|||||
DB 268 TGTGGAAATTTTCCTTTGATA 289

RESULT 11
AV546877
LOCUS      AV546877      359 bp      mRNA      EST      07-SEP-2000
DEFINITION      AV546877 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
CDNA clone R2L21a12F 3', mRNA sequence.
ACCESSION      AV546877

```

```

VERSION  AV546877.1  GI:8718291
KEYWORDS  EST.
SOURCE    Arabidopsis thaliana
ORGANISM  thale cress.
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS  Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
TITLE    Brassicales; Brassicaceae; Arabidopsis.
JOURNAL  1 (bases 1 to 359)
MEDLINE  Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
COMMENT  A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
    source
        Location/Qualifiers
            1..359
                /organism="Arabidopsis thaliana"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /clone_lib="RZL21a12f"
                /clone_lib="Arabidopsis thaliana roots Columbia"
                /tissue_type="roots"
                /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                XhoI"
BASE COUNT  107 a 65 c 47 g 140 t
ORIGIN
Query Match      63.6%; Score 14; DB 37; Length 359;
Best Local Similarity 63.6%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY  1  tgttgannnnnnntttgata 22
      |||||
      |||||
Db   8  TGTGGAACTATTATTGATA 29

RESULT 12
AW226052
LOCUS     ST7608 Pine Triplex shoot tip library Pinus taeda cDNA clone
DEFINITION ST7608, mRNA sequence.
ACCESSION AW226052
VERSION   AW226052.1  GI:6555348
KEYWORDS  EST.
SOURCE    loblolly pine.
ORGANISM  Pinus taeda
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS  Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
TITLE    1 (bases 1 to 364)
JOURNAL  Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
COMMENT  The Pine Gene Discovery Project
Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 sequencing Primer.
Location/Qualifiers
    1..364
        /organism="Pinus taeda"
        /db_xref="taxon:3352"
        /clone="ST7608"
        /clone_lib="Pine Triplex shoot tip library"

```

```

/lab_host="E. coli BM25.8"
/note="Organ: shoot tips; Vector: Lambda Triplex; Site_1:
SfiI (A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
Lambda triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."
BASE COUNT  106 a 48 c 74 g 136 t
ORIGIN
Query Match      63.6%; Score 14; DB 87; Length 364;
Best Local Similarity 63.6%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY  1  tgttgannnnnnntttgata 22
      |||||
      |||||
Db   99  TGTGGATGAAGTTTTCGATA 120

RESULT 13
AQ311167/c
LOCUS     CITBI-E1-2525E11.TR CITBI-E1 Homo sapiens genomic clone 2525E11,
DEFINITION DNA sequence.
ACCESSION AQ311167
VERSION   AQ311167.1  GI:4042980
KEYWORDS  GSS.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 371)
          Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
          Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
          Venter, J.C.
          Use of a random human BAC End Sequence Database for Sequence-Ready
          Map Building
          Unpublished (1998)
          Contact: Shaying Zhao, William Nierman, Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: hbe@tigr.org
          Clones are available from Research Genetics (info@resgen.com). BAC
          end search page:
          http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
          Seq primer: M13 Reverse
          Class: BAC ends.
FEATURES
    source
        Location/Qualifiers
            1..371
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="2525E11"
                /clone_lib="CITBI-E1"
                /sex="male"
                /cell_type="sperm"
                /note="Vector: pBelBAC11; Site_1: EcoRI; Site_2: EcoRI;
                Caltech Human BAC Library D"
BASE COUNT  131 a 69 c 85 g 86 t
ORIGIN
Query Match      63.6%; Score 14; DB 151; Length 371;
Best Local Similarity 63.6%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY  1  tgttgannnnnnntttgata 22
      |||||
      |||||

```

Db 276 TGTGGAAATTACAGTTTGATA 255

RESULT 14  
AQ102282/c

LOCUS  
DEFINITION

AQ102282 376 bp DNA GSS 27-AUG-1998  
HS\_3029\_A2\_G04\_MF CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=3029 Col=8 Row=M, DNA sequence.

ACCESSION  
AQ102282

VERSION  
AQ102282.1 GI:3473311

KEYWORDS  
GSS.

SOURCE  
human.

ORGANISM  
Homo sapiens

REFERENCE  
1 (bases 1 to 376)

AUTHORS  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

TITLE  
Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

JOURNAL  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE  
99380589

COMMENT  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3029 row: M column: 8  
Class: BAC ends  
High quality sequence stop: 376.

FEATURES

Source

BASE COUNT  
ORIGIN

131 a 81 c 50 g 114 t

Query Match

Best Local Similarity 63.6%; Score 14; DB 149; Length 376;

Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22

|||||

Db 333 TGTGGAAATTGCCAATTGATA 312

RESULT 15  
AQ844476/c

LOCUS  
DEFINITION

AQ844476 376 bp DNA GSS 03-NOV-1999  
an37g11 JM101 filtered library zea mays genomic, DNA sequence.

ACCESSION  
AQ844476

VERSION  
AQ844476.1 GI:6202964

KEYWORDS  
GSS.

SOURCE  
Zea mays.

ORGANISM  
Zea mays

REFERENCE  
1 (bases 1 to 376)

AUTHORS  
Rabinowicz,P.D., Schutz,K., Dedhia,N., Yordan,C., Parnell,L.D.,  
Stein,L., McCombie,W.R. and Martienssen,R.A.

TITLE  
Differential methylation of genes and retrotransposons allows

shotgun sequencing of the maize genome

Nat. Genet. 23, 305-308 (1999)

COMMENT

Contact: Martienssen RA  
Cold Spring Harbor Laboratory  
1 Bungtown Rd., Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8322  
Fax: 516 367 8369  
Email: martienssen@cshl.org  
Seq primer: forward  
Class: shotgun.

FEATURES

Source

Location/Qualifiers

1..376

/organism="Zea mays"

/cultivar="B73"

/db\_xref="taxon:4577"

/clone\_lib="JM101 filtered library"

/note="Organ: Immature ears; Vector: M13; Site\_1: Xba I;

DNA prepared from purified nuclei was digested with the

methylation insensitive enzyme Spe I, size fractionated to

enrich for the 0.5 to 4 kbp fraction, ligated into Xba I

digested M13 vector and electroporated into E.coli JM101.

BASE COUNT 89 a 87 c 100 g 100 t

ORIGIN

Query Match

Best Local Similarity 63.6%; Score 14; DB 163; Length 376;

Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tgttgannnnnnntttgata 22

|||||

Db 109 TGTGGATCTGGAAGTTTGATA 88

Search completed: March 27, 2001, 07:38:25

Job time: 4588 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:23:01 ; Search time 207.51 Seconds  
(without alignments)  
39.827 Million cell updates/sec

Title: US-09-380-826A-7  
Perfect score: 22  
Sequence: 1 tgggtgacacagatttgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

1: /cgn2\_2/gcgdata/geneseq/geneseq/NA1980.DAT:\*

2: /cgn2\_2/gcgdata/geneseq/geneseq/NA1981.DAT:\*

3: /cgn2\_2/gcgdata/geneseq/geneseq/NA1982.DAT:\*

4: /cgn2\_2/gcgdata/geneseq/geneseq/NA1983.DAT:\*

5: /cgn2\_2/gcgdata/geneseq/geneseq/NA1984.DAT:\*

6: /cgn2\_2/gcgdata/geneseq/geneseq/NA1985.DAT:\*

7: /cgn2\_2/gcgdata/geneseq/geneseq/NA1986.DAT:\*

8: /cgn2\_2/gcgdata/geneseq/geneseq/NA1987.DAT:\*

9: /cgn2\_2/gcgdata/geneseq/geneseq/NA1988.DAT:\*

10: /cgn2\_2/gcgdata/geneseq/geneseq/NA1989.DAT:\*

11: /cgn2\_2/gcgdata/geneseq/geneseq/NA1990.DAT:\*

12: /cgn2\_2/gcgdata/geneseq/geneseq/NA1991.DAT:\*

13: /cgn2\_2/gcgdata/geneseq/geneseq/NA1992.DAT:\*

14: /cgn2\_2/gcgdata/geneseq/geneseq/NA1993.DAT:\*

15: /cgn2\_2/gcgdata/geneseq/geneseq/NA1994.DAT:\*

16: /cgn2\_2/gcgdata/geneseq/geneseq/NA1995.DAT:\*

17: /cgn2\_2/gcgdata/geneseq/geneseq/NA1996.DAT:\*

18: /cgn2\_2/gcgdata/geneseq/geneseq/NA1997.DAT:\*

19: /cgn2\_2/gcgdata/geneseq/geneseq/NA1998.DAT:\*

20: /cgn2\_2/gcgdata/geneseq/geneseq/NA1999.DAT:\*

21: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	19 V58901	Leptospiira rRNA ge
2	22	100.0	22	19 V58897	L. fairnei nucleoti
3	22	100.0	1477	19 V58896	L. fairnei nucleoti
C 4	16.8	76.4	2582	18 T90543	Maize male fertili
C 5	16.2	73.6	951	20 V84598	Human secreted pro
C 6	16.2	73.6	1291	19 V59636	Human secreted pro
C 7	16.2	73.6	1473	20 X61757	B. burgdorferi ant
8	16.2	73.6	2000	17 T17111	Acetyl-CoA-carboxy
9	16.2	73.6	2001	17 T39904	Maize acetyl CoA c
10	16.2	73.6	2001	21 T49816	ECORI fragment of
11	16.2	73.6	3822	18 T64683	M. leprae gyra pre
12	16.2	73.6	4346	14 QN2933	A3 maize ACCase CD

13	16.2	73.6	7470	17 T39905	Maize acetyl CoA c
14	16.2	73.6	7470	19 V29317	Maize ACCase enzym
15	16.2	73.6	7470	21 Z49820	Maize acetyl CoA c
C 16	16.2	73.6	10811	19 V59091	Potato pollen cell
17	15.8	71.8	843	20 Z15467	Human gene express
18	15.8	71.8	961	19 X14095	H. pylori GHPO 131
19	15.8	71.8	4146	21 Z51556	Human hypoxia resp
C 20	15.6	70.9	297	20 V88753	EST clone HK630.
C 21	15.6	70.9	795	20 V72025	Adenovirus PACTSG2
C 22	15.6	70.9	834	20 V72026	Adenovirus SCAR.RG
C 23	15.6	70.9	1075	18 T72715	C. elegans inhibit
C 24	15.6	70.9	1096	13 Q20498	Encodes fibrinogen
C 25	15.6	70.9	1194	20 V72027	Adenovirus PACSG2S
C 26	15.6	70.9	1560	18 T49864	Callosellasma rhod
C 27	15.6	70.9	1584	19 V28845	Human coxsackievir
C 28	15.6	70.9	1761	21 Z45339	DNA encoding a Bac
C 29	15.6	70.9	1989	20 V81481	Black Creek Canal
C 30	15.6	70.9	2354	19 V50429	Human coxsackievir
C 31	15.6	70.9	2851	20 X13369	Enterococcus faeca
C 32	15.6	70.9	7745	18 T79859	Genomic DNA encodi
C 33	15.6	70.9	7745	18 T59308	Sh2-mlRev6 gene (s
C 34	15.6	70.9	7745	20 X04183	Shrunken-2 gene va
C 35	15.6	70.9	9824	14 Q40763	Wild-type shrunken
C 36	15.2	69.1	349	13 Q34809	Human type IV coll
C 37	15.2	69.1	493	20 X13759	Enterococcus faeca
C 38	15.2	69.1	1028	17 T36923	Human OVCA2 tumou
C 39	15.2	69.1	1848	16 Q86996	Aminopeptidase O12
C 40	15.2	69.1	1848	16 Q86998	Aminopeptidase O12
C 41	15.2	69.1	2202	17 T36922	Human OVCA1 tumou
C 42	15.2	69.1	2361	20 X04316	Human secreted pro
C 43	15.2	69.1	2598	18 T72286	Breast cancer mann
C 44	15.2	69.1	2598	21 Z98878	MMTV env protein n
C 45	15.2	69.1	5420	12 Q11643	Partial human comp

ALIGNMENTS

RESULT 1

V58901

ID V58901 standard; DNA; 22 BP.

XX

AC V58901;

XX

DT 20-JAN-1999 (first entry)

XX

DE Leptospiira rRNA gene nucleotide sequence.

XX

KW Infection; pathogenic Leptospiira; protective immunity; therapy;

KW diagnosis; ss.

XX

OS Leptospiira sp.

PN WO9840099-A1.

XX

PD 17-SEP-1998.

XX

PF 06-MAR-1998; 98WO-AU00145.

XX

PR 07-MAR-1997; 97AU-0005494.

XX

PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

PA (PIGR-) PIG RES & DEV CORP.

XX

PI Chappel RJ;

XX

XX WPI; 1998-520791/44.

XX

PT New isolated pathogenic Leptospiira bacterium - useful for, e.g

PT developing products for conferring protective immunity, and for

PT prophylactic or therapeutic treatment

XX

PS Claim 15; Page 72; 94pp; English.

XX This sequence represents a Leptospira DNA sequence isolated from the  
 CC pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 CC belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 CC L. fainei. The LS bacteria can be used for conferring protective  
 CC immunity against pathogenic LS bacteria in humans or animals. The  
 CC bacteria can also be used for prophylactic or therapeutic treatment of LS  
 CC infections. The DNAs and antibodies may also be used for detection and  
 CC diagnosis of past or present LS infection.  
 XX  
 SQ Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.047;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaatttgata 22  
 |||||  
 Db 1 tgttgatcacagaatttgata 22

RESULT 2  
 V58897  
 ID V58897 standard; DNA; 22 BP.  
 AC  
 XX  
 AC V58897;

20-JAN-1999 (first entry)

L. fainei nucleotide sequence.

Infection; pathogenic Leptospira; protective immunity; therapy;  
 diagnosis; ss.

Leptospira fainei.

WO9840099-A1.

17-SEP-1998.

06-MAR-1998; 98WO-AU00145.

07-MAR-1997; 97AU-0005494.

(AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 (PIGR-) PIG RES & DEV CORP.

Chappel RJ;

WPI; 1998-520791/44.

New isolated pathogenic Leptospira bacterium - useful for, e.g  
 developing products for conferring protective immunity, and for  
 prophylactic or therapeutic treatment

Claim 15; Page 70; 94pp; English.

This sequence represents a Leptospira DNA sequence isolated from the  
 pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 L. fainei. The LS bacteria can be used for conferring protective  
 immunity against pathogenic LS bacteria in humans or animals. The  
 bacteria can also be used for prophylactic or therapeutic treatment of LS  
 infections. The DNAs and antibodies may also be used for detection and  
 diagnosis of past or present LS infection.

Sequence 22 BP; 7 A; 2 C; 5 G; 8 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.047;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaatttgata 22  
 |||||  
 Db 1 tgttgatcacagaatttgata 22

RESULT 3

V58896

ID V58896 standard; DNA; 1477 BP.

AC V58896;

20-JAN-1999 (first entry)

L. fainei nucleotide sequence.

Infection; pathogenic Leptospira; protective immunity; therapy;  
 diagnosis; ss.

Leptospira fainei.

WO9840099-A1.

17-SEP-1998.

06-MAR-1998; 98WO-AU00145.

07-MAR-1997; 97AU-0005494.

(AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 (PIGR-) PIG RES & DEV CORP.

Chappel RJ;

WPI; 1998-520791/44.

New isolated pathogenic Leptospira bacterium - useful for, e.g  
 developing products for conferring protective immunity, and for  
 prophylactic or therapeutic treatment

Claim 15; Page 69-70; 94pp; English.

This sequence represents a Leptospira DNA sequence isolated from the  
 pathogenic Leptospira (LS) bacterium of the invention. The bacterium  
 belongs to serogroup Hurstbridge or serovar hurstbridge or the species  
 L. fainei. The LS bacteria can be used for conferring protective  
 immunity against pathogenic LS bacteria in humans or animals. The  
 bacteria can also be used for prophylactic or therapeutic treatment of LS  
 infections. The DNAs and antibodies may also be used for detection and  
 diagnosis of past or present LS infection.

Sequence 1477 BP; 390 A; 334 C; 439 G; 314 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 1477;  
 Best Local Similarity 100.0%; Pred. No. 0.072;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgttgatcacagaatttgata 22  
 |||||  
 Db 154 tgttgatcacagaatttgata 175

RESULT 4

T90543/C

ID T90543 standard; DNA; 2582 BP.

AC T90543;

13-FEB-1998 (first entry)

Maize male fertility gene 233 Zm41-A.

KW Zm41-A; male fertility; hybrid seed; transgenic plant; antisense;  
 KW ribozyme; male sterile; maize; Ms41-A; ds.  
 XX

OS Zea mays.

FH Key Location/Qualifiers

FT exon 823..915

FT intron /\*tag= a

FT exon 916..935

FT exon /\*tag= b

FT exon 936..1090

FT intron /\*tag= c

FT exon 1091..1167

FT exon /\*tag= d

FT exon 1168..1332

FT exon /\*tag= e

XX WO9723618-A1.

XX PN

XX PD

XX PF

XX 20-DEC-1996; 96WO-GB03191.

XX 21-DEC-1995; 95GB-0026218.

XX (GENE-) GENE SHEARS PTY LTD.

XX Baudot G, Garcia D, Hodge R, Perez P;

XX WPI; 1997-351055/32.

XX Nucleic acid encoding proteins involved in male fertility in plants

XX - used to control fertility and for production of hybrid seed

XX Claim 3; Fig 15; 85pp; English.

XX This DNA sequence comprises the maize Z33 Zm41-A gene. This is an

XX orthologue of Arabidopsis Ms41-A (see T90522), a gene that confers

XX male fertility. It was obtained following a database search for

XX sequences that showed homology to Ms41-A DNA, isolation of a

XX partial cDNA clone, and use of this clone to isolate Zm41-A genes

XX Z31 (T90542), Z33 (T90543) and Z35 (T90544) from genomic lambda

XX libraries. The genes show a high level of conservation. Z35 may

XX be derived from Z31 via genetic rearrangements, deletions and/or

XX insertions. Z33 has subsequent deletions from Z35 and is

XX truncated, having only exons 3, 5 and 6. The Ms41-A and Zm41-A

XX genes, antisense or ribozyme sequences can be used to produce

XX transgenic plants with controlled male fertility. Male sterile

XX plants are useful for hybrid seed production, particularly in

XX plants where restoration of fertility is not needed, e.g.

XX Brassicaceae, lettuce, spinach and onions.

XX

XX Sequence 2582 BP; 670 A; 512 C; 537 G; 863 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 XX Homo sapiens.  
 XX WO9854963-A2.  
 XX 10-DEC-1998.  
 XX 04-JUN-1998; 98WO-US11422.  
 XX 18-DEC-1997; 97US-0070923.  
 XX 06-JUN-1997; 97US-0048877.  
 XX 06-JUN-1997; 97US-0048881.  
 XX 06-JUN-1997; 97US-0048884.  
 XX 06-JUN-1997; 97US-0048893.  
 XX 06-JUN-1997; 97US-0048896.  
 XX 06-JUN-1997; 97US-0048899.  
 XX 06-JUN-1997; 97US-0048915.  
 XX 06-JUN-1997; 97US-0048949.  
 XX 06-JUN-1997; 97US-0048964.  
 XX 06-JUN-1997; 97US-0048972.  
 XX 06-JUN-1997; 97US-0049020.  
 XX 06-JUN-1997; 97US-0049375.  
 XX 05-SEP-1997; 97US-0057628.  
 XX 05-SEP-1997; 97US-0057635.  
 XX 05-SEP-1997; 97US-0057644.  
 XX 05-SEP-1997; 97US-0057647.  
 XX 05-SEP-1997; 97US-0057650.  
 XX 05-SEP-1997; 97US-0057661.  
 XX 05-SEP-1997; 97US-0057667.  
 XX 05-SEP-1997; 97US-0057761.  
 XX 05-SEP-1997; 97US-0057764.  
 XX 05-SEP-1997; 97US-0057770.  
 XX 05-SEP-1997; 97US-0057775.  
 XX 05-SEP-1997; 97US-0057778.  
 XX 06-JUN-1997; 97US-0048875.  
 XX 06-JUN-1997; 97US-0048878.  
 XX 06-JUN-1997; 97US-0048882.  
 XX 06-JUN-1997; 97US-0048885.  
 XX 06-JUN-1997; 97US-0048894.  
 XX 06-JUN-1997; 97US-0048897.  
 XX 06-JUN-1997; 97US-0048900.  
 XX 06-JUN-1997; 97US-0048916.  
 XX 06-JUN-1997; 97US-0048962.  
 XX 06-JUN-1997; 97US-0048970.  
 XX 06-JUN-1997; 97US-0048974.  
 XX 06-JUN-1997; 97US-0049373.  
 XX 05-SEP-1997; 97US-0057584.  
 XX 05-SEP-1997; 97US-0057629.  
 XX 05-SEP-1997; 97US-0057642.  
 XX 05-SEP-1997; 97US-0057645.  
 XX 05-SEP-1997; 97US-0057648.  
 XX 05-SEP-1997; 97US-0057651.  
 XX 05-SEP-1997; 97US-0057662.  
 XX 05-SEP-1997; 97US-0057668.  
 XX 05-SEP-1997; 97US-0057762.  
 XX 05-SEP-1997; 97US-0057765.  
 XX 05-SEP-1997; 97US-0057771.  
 XX 06-JUN-1997; 97US-0048876.  
 XX 06-JUN-1997; 97US-0048880.  
 XX 06-JUN-1997; 97US-0048883.  
 XX 06-JUN-1997; 97US-0048892.  
 XX 06-JUN-1997; 97US-0048895.  
 XX 06-JUN-1997; 97US-0048898.  
 XX 06-JUN-1997; 97US-0048899.

```

PR 06-JUN-1997; 97US-0048901.
PR 06-JUN-1997; 97US-0048917.
PR 06-JUN-1997; 97US-0048963.
PR 06-JUN-1997; 97US-0048971.
PR 06-JUN-1997; 97US-0049019.
PR 06-JUN-1997; 97US-0049374.
PR 05-SEP-1997; 97US-0057627.
PR 05-SEP-1997; 97US-0057634.
PR 05-SEP-1997; 97US-0057643.
PR 05-SEP-1997; 97US-0057646.
PR 05-SEP-1997; 97US-0057649.
PR 05-SEP-1997; 97US-0057654.
PR 05-SEP-1997; 97US-0057666.
PR 05-SEP-1997; 97US-0057760.
PR 05-SEP-1997; 97US-0057763.
PR 05-SEP-1997; 97US-0057769.
PR 05-SEP-1997; 97US-0057774.
PR 05-SEP-1997; 97US-0057777.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
XX Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;
XX Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;
XX Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
XX Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
XX
XX WPI; 1999-059865/05.
XX P-PSDB; W88721.
XX
XX New isolated human genes and the secreted polypeptides they encode -
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 4; Page 450-451; 772pp; English.
XX
XX The invention relates to nucleic acid sequences (V84411 to V84633)
XX encoding human secreted proteins (W88534 to W88756). The secreted protein
XX gene sequences are deposited with the ATCC under deposit numbers AFCC
XX 97979, 97974, 97975, 97976, 97977, 209007, 209009, 209010,
XX 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
XX cells comprising recombinant vectors containing the nucleic acid
XX sequences are used for the recombinant production of the secreted
XX proteins. The polynucleotide and amino acid sequences are useful for are
XX useful for preventing, treating or ameliorating medical conditions e.g.
XX by protein or gene therapy. Pathological conditions can be also diagnosed
XX by determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the polynucleotides, based on
XX which tissues they are most highly expressed in, and include developing
XX products for the diagnosis or treatment of cancer, neurodegenerative
XX disorders, developmental abnormalities and foetal deficiencies, blood
XX disorders, tumours, leukemias, diseases of the immune system, autoimmune
XX diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
XX ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
XX restenosis, prostate diseases, obesity, disorders involving osteoclasts
XX such as osteoporosis, arthritis or malignancies, diseases of testes, lung
XX or thymus, digestive/endocrine disorders, infections and AIDS. The
XX polypeptides are also useful for identifying their binding partners.
XX The present sequence represents a gene encoding a human secreted protein
XX (see descriptor line for gene number and clone identification).
XX
XX Sequence 951 BP; 296 A; 134 C; 192 G; 329 T; 0 other;
XX
XX Query Match 73.6%; Score 16.2; DB 20; Length 951;
XX Best Local Similarity 85.7%; Pred. No. 47;
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX Qy 1 tggtagatcacaaagatttgat 21
XX Db 524 tggtagatcacaaagatttgat 504

```

```

RESULT 6
V59636/C
ID V59636 standard; DNA; 1291 BP.
XX
XX AC V59636;
XX
XX DT 19-JAN-1999 (first entry)
XX
XX DE Human secreted protein gene 126 clone HELGH31.
XX
XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
XX immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
XX inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
XX osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
XX OS Homo sapiens.
XX
XX PN W09839448-A2.
XX
XX PD 11-SEP-1998.
XX
XX PF 06-MAR-1998; 98WO-US04493.
XX
XX PR 02-OCT-1997; 97US-0061060.
XX PR 07-MAR-1997; 97US-0038621.
XX PR 07-MAR-1997; 97US-0040161.
XX PR 07-MAR-1997; 97US-0040162.
XX PR 07-MAR-1997; 97US-0040163.
XX PR 07-MAR-1997; 97US-0040333.
XX PR 07-MAR-1997; 97US-0040334.
XX PR 07-MAR-1997; 97US-0040336.
XX PR 07-MAR-1997; 97US-0040626.
XX PR 11-APR-1997; 97US-0043311.
XX PR 11-APR-1997; 97US-0043312.
XX PR 11-APR-1997; 97US-0043313.
XX PR 11-APR-1997; 97US-0043314.
XX PR 11-APR-1997; 97US-0043568.
XX PR 11-APR-1997; 97US-0043569.
XX PR 11-APR-1997; 97US-0043576.
XX PR 11-APR-1997; 97US-0043578.
XX PR 11-APR-1997; 97US-0043580.
XX PR 11-APR-1997; 97US-0043669.
XX PR 11-APR-1997; 97US-0043670.
XX PR 11-APR-1997; 97US-0043671.
XX PR 11-APR-1997; 97US-0043672.
XX PR 11-APR-1997; 97US-0043674.
XX PR 23-MAY-1997; 97US-0047492.
XX PR 23-MAY-1997; 97US-0047500.
XX PR 23-MAY-1997; 97US-0047501.
XX PR 23-MAY-1997; 97US-0047502.
XX PR 23-MAY-1997; 97US-0047503.
XX PR 23-MAY-1997; 97US-0047581.
XX PR 23-MAY-1997; 97US-0047582.
XX PR 23-MAY-1997; 97US-0047583.
XX PR 23-MAY-1997; 97US-0047584.
XX PR 23-MAY-1997; 97US-0047585.
XX PR 23-MAY-1997; 97US-0047586.
XX PR 23-MAY-1997; 97US-0047587.
XX PR 23-MAY-1997; 97US-0047588.
XX PR 23-MAY-1997; 97US-0047589.
XX PR 23-MAY-1997; 97US-0047590.
XX PR 23-MAY-1997; 97US-0047592.
XX PR 23-MAY-1997; 97US-0047593.
XX PR 23-MAY-1997; 97US-0047594.
XX PR 23-MAY-1997; 97US-0047595.
XX PR 23-MAY-1997; 97US-0047596.
XX PR 23-MAY-1997; 97US-0047597.
XX PR 23-MAY-1997; 97US-0047598.

```



23-MAY-1997; 97US-0047599.  
 23-MAY-1997; 97US-0047600.  
 23-MAY-1997; 97US-0047601.  
 23-MAY-1997; 97US-0047612.  
 23-MAY-1997; 97US-0047613.  
 23-MAY-1997; 97US-0047614.  
 23-MAY-1997; 97US-0047615.  
 23-MAY-1997; 97US-0047617.  
 23-MAY-1997; 97US-0047618.  
 23-MAY-1997; 97US-0047632.  
 23-MAY-1997; 97US-0047633.  
 06-JUN-1997; 97US-0048964.  
 06-JUN-1997; 97US-0048974.  
 13-JUN-1997; 97US-0049610.  
 08-JUL-1997; 97US-0051926.  
 16-JUL-1997; 97US-0052874.  
 18-AUG-1997; 97US-0055724.  
 22-AUG-1997; 97US-0056630.  
 22-AUG-1997; 97US-0056631.  
 22-AUG-1997; 97US-0056632.  
 22-AUG-1997; 97US-0056636.  
 22-AUG-1997; 97US-0056637.  
 22-AUG-1997; 97US-0056662.  
 22-AUG-1997; 97US-0056664.  
 22-AUG-1997; 97US-0056845.  
 22-AUG-1997; 97US-0056862.  
 22-AUG-1997; 97US-0056864.  
 22-AUG-1997; 97US-0056872.  
 22-AUG-1997; 97US-0056874.  
 22-AUG-1997; 97US-0056875.  
 22-AUG-1997; 97US-0056876.  
 22-AUG-1997; 97US-0056877.  
 22-AUG-1997; 97US-0056878.  
 22-AUG-1997; 97US-0056879.  
 22-AUG-1997; 97US-0056880.  
 22-AUG-1997; 97US-0056881.  
 22-AUG-1997; 97US-0056882.  
 22-AUG-1997; 97US-0056884.  
 22-AUG-1997; 97US-0056886.  
 22-AUG-1997; 97US-0056887.  
 22-AUG-1997; 97US-0056888.  
 22-AUG-1997; 97US-0056889.  
 22-AUG-1997; 97US-0056892.  
 22-AUG-1997; 97US-0056893.  
 22-AUG-1997; 97US-0056894.  
 22-AUG-1997; 97US-0056903.  
 22-AUG-1997; 97US-0056908.  
 22-AUG-1997; 97US-0056909.  
 22-AUG-1997; 97US-0056910.  
 22-AUG-1997; 97US-0056911.  
 05-SEP-1997; 97US-0057650.  
 05-SEP-1997; 97US-0057669.  
 05-SEP-1997; 97US-0057761.  
 12-SEP-1997; 97US-0058785.  
 (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;  
 PI Feng P, Ferris AM, Fischer CL, Florence KA, Greene JM, Hu JS;  
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 XX WPI; 1998-506364/43.  
 DR P-PSDB; W74854.  
 XX  
 XX New isolated human genes and the secreted polypeptide(s) they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 XX  
 PS Claim 1; Page 360-361; 721pp; English.  
 XX  
 XX This sequence represents a nucleic acid molecule designated Gene 126 from  
 CC the human cDNA clone HELGH31 (deposited as clone ATCC 97902 and ATCC

209048) which encodes a secreted human protein. The gene can be used to  
 generate fusion proteins by linking to the gene to a human immunoglobulin  
 Fc portion (e.g. V59502) for increasing the stability of the fused  
 protein as compared to the human protein only.  
 CC The invention relates to 186 novel genes and their fragments (nucleic  
 CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also, pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the 186 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see V59511 for described  
 CC uses).  
 XX  
 XX Sequence 1291 BP; 391 A; 228 C; 203 G; 469 T; 0 other;  
 SQ

Query Match 73.6%; Score 16.2; DB 19; Length 1291;  
 Best Local Similarity 85.7%; Pred. No. 48;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 tgttgatcacagattgat 21  
 Db 367 TGTATGATCAGATTGAT 347  
 III IIIIIII IIIIIII  
 RESULT 7  
 X61757/c  
 ID X61757 standard; DNA; 1473 BP.  
 XX  
 AC X61757;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein coding sequence, f893.nt.  
 XX  
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.  
 XX  
 OS Borrelia burgdorferi.  
 XX  
 PN WO9859071-A1.  
 XX  
 XX 30-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-US12718.  
 XX  
 PR 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 XX  
 XX Choi GH, Erwin AL, Hanson MS, Lathigra R;  
 XX  
 DR WPI; 1999-189980/16.  
 DR P-PSDB; Y20060.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by Borrelia, particularly Lyme disease  
 XX  
 XX Claim 1; Page 180; 275pp; English.  
 XX  
 XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.

1.

QY 2 gttggatcacagaatttgata 22  
 ID T64683 standard; DNA; 2001 BP.  
 DB 42 gttggatcacagaatttgata 62

RESULT 10  
 249816  
 ID 249816 standard; DNA; 2001 BP.  
 AC 249816;  
 XX  
 DT 18-APR-2000 (first entry)  
 DE EcoRI fragment of lambda clone #15-14 with maize ACCase gene portion.  
 KW Herbicide resistance; gene modification; lambda clone #15-14;  
 KW maize acetyl CoA carboxylase; ACCase; herbicide; plant oil content;  
 KW marker-assisted plant selection; groat oil trait;  
 KW restriction fragment length polymorphism mapping;  
 KW high-energy animal feed; low-fat human food; ds.  
 XX  
 OS Zea mays.  
 XX  
 PN WO9967367-Al.  
 XX  
 PD 29-DEC-1999.  
 XX  
 PF 22-JUN-1999; 99WO-US14022.  
 XX  
 PR 22-JUN-1998; 98US-0090240.  
 PR 02-JUL-1998; 98US-0091640.  
 XX  
 PA (MINU ) UNIV MINNESOTA.  
 PA (USDA ) US DEPT OF AGRICULTURE.  
 PA (EGLI/) EGLI M A.  
 PA (GROH/) GROH S.  
 PA (KIAN/) KIANIAN S F.  
 PA (PHIL/) PHILLIPS R L.  
 PA (RINE/) RINES H W.  
 PA (SOME/) SOMERS D A.  
 XX  
 PI Egli MA, Groh S, Kianian SF, Phillips RL, Rines HW, Somers DA;  
 XX WPI: 2000-147205/13.  
 DR  
 XX  
 PT New DNA encoding acetyl-CoA carboxylase from oats, used to produce  
 PT transformed plants with herbicide resistance and altered oil content  
 PS Disclosure; Fig 10; 197pp; English.  
 XX  
 CC The present DNA sequence is a 2kb EcoRI fragment of lambda clone #15-14  
 CC including a portion of a maize acetyl CoA carboxylase (ACCase) gene  
 CC located at bases 2883 to 83 from the 3' stop codon. ACCase is involved  
 CC in fatty acid synthesis and is the target of the specified herbicides.  
 CC Transformation of plants with ACCase imparts resistance to  
 CC cyclohexanedione and aryloxyphenoxypropionic acid herbicides and alter  
 CC the oil content. The ACCase DNA is also used as source of probes and  
 CC primers for the identification of transgenic plants; in marker-assisted  
 CC plant selection and for restriction fragment length polymorphism  
 CC mapping, used for high-energy animal feed and high-fiber, low-fat human  
 CC food and in genetic dissection of the groat oil trait.  
 XX  
 SQ Sequence 2001 BP; 546 A; 406 C; 494 G; 554 T; 1 other;

Query Match 73.6%; Score 16.2; DB 21; Length 2001;  
 Best Local Similarity 85.7%; Pred. No. 51;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggatcacagaatttgata 22  
 ID T64683 standard; DNA; 2001 BP.  
 DB 42 gttggatcacagaatttgata 62

RESULT 11  
 T64683  
 ID T64683 standard; DNA; 3822 BP.  
 XX  
 AC T64683;  
 XX  
 DT 17-MAR-1998 (first entry)  
 DE  
 XX  
 M. leprae gyra precursor coding sequence.  
 XX  
 KW Mycobacterium sp.; internal sequence; intein; immature; gyrase;  
 KW protein splicing; precursor; gyra; ss.  
 XX  
 OS Mycobacterium leprae.  
 XX  
 PN FR2739859-Al.  
 XX  
 PD 18-APR-1997.  
 XX  
 PF 17-OCT-1995; 95PR-0012162.  
 XX  
 PR 17-OCT-1995; 95PR-0012162.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Cole S, Fsihi H;  
 XX  
 DR WPI: 1997-247976/23.  
 DR P-PSDB; W15078.  
 XX  
 PT Mycobacterial DNA gyrase precursor protein - and polypeptide(s)  
 PT corresponding to mycobacterial DNA gyrase intein sequences  
 XX  
 PS Claim 5; Pages 33-34; 55pp; French.  
 XX  
 CC This DNA, isolated from Mycobacterium leprae, encodes a precursor  
 CC (immature) gyrase protein from which an "intein" (see W15074) is excised  
 CC during maturation. The gyrase and its coding sequence are useful to  
 CC fabricate products that alter the maturation of proteins essential for  
 CC the development of infectious agents by altering the protein splicing of  
 CC precursor polypeptides of the proteins. The inteins encoded by the gyra  
 CC genes of Mycobacterium leprae, M. flavesens, M. goodii and M. kansasii  
 CC are used: (a) to modify the genome of a eukaryotic cell that lacks  
 CC endogenous biological activity identical to that of the polypeptide;  
 CC (b) to replace a copy of a gene present in a recipient genome by  
 CC integration of a gene different from that where the integration takes  
 CC place; and (c) for targeted insertion of a foreign DNA sequence into a  
 CC selected site in the genome of a eukaryotic cell not containing the  
 CC specific cleavage sites of the polypeptide.  
 XX  
 SQ Sequence 3822 BP; 843 A; 899 C; 1092 G; 988 T; 0 other;

Query Match 73.6%; Score 16.2; DB 18; Length 3822;  
 Best Local Similarity 85.7%; Pred. No. 54;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggatcacagaatttgata 22  
 ID T64683 standard; DNA; 3822 BP.  
 DB 2334 gttggatcacagaatttgata 2354

RESULT 12  
 Q42933  
 ID Q42933 standard; DNA; 4346 BP.  
 XX  
 AC Q42933;  
 XX  
 DT 15-OCT-1993 (first entry)  
 XX  
 DE A3 maize ACCase cDNA clone.

XX Acetyl CoA carboxylase; ACCase; probe; expression; oil content;  
 KW lower; modified; increased; oilseed rape; sunflower; resistant;  
 KW resistance; grass-weed herbicides; fluzafop; alkylketones; maize;  
 KW wheat; barley; ss.  
 XX  
 OS Zea mays.  
 XX  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 3..3947  
 FT /\*tag= a  
 FT /note= "ACCase"  
 FT misc\_feature 1875..1878  
 FT /\*tag= b  
 FT /note= "GAN encodes Glu or Asp"  
 XX  
 PN W09311243-A.  
 XX  
 XX  
 PD 10-JUN-1993.  
 XX  
 XX  
 PF 27-NOV-1992; 92WO-CB02205.  
 XX  
 XX  
 PR 28-NOV-1991; 91GB-0025330.  
 XX  
 XX (ICIL ) ICI AUSTRALIA OPERATIONS PTY LTD.  
 PA (ICIL ) IMPERIAL CHEM IND PLC.  
 XX  
 XX Ashton AR, Jenkins CL, Whitfield PR;  
 PI  
 XX WPI: 1993-197061/24.  
 DR P-PSDB; R36781.  
 XX  
 XX DNA clones for use in probing plant DNA - comprise maize acetyl  
 PT coenzyme-A carboxylase gene flanked by heterologous DNA  
 PT  
 XX  
 XX Claim 1; Fig 3; 69pp; English.  
 PS  
 XX The sequence is that of the A3 acetyl CoA carboxylase (ACCase) cDNA  
 CC maize clone which can be used to probe plant DNA to isolate other  
 CC such clones. It may be used with plant regulatory sequences to  
 CC produce expression cassettes for ACCase. These cassettes are used to  
 CC transform plants to downregulate prodn. of ACCase. This can alter  
 CC the composition of seeds or other plant parts, e.g. enabling oil-  
 CC bearing plants such as oilseed rape, sunflower or oilpalm, having a  
 CC lower or modified oil content to be produced. It can be used to form  
 CC expression cassettes for overexpression of ACCase., leading to prodn.  
 CC of plants with an increased oil content. It may also be used to recover  
 CC the ACCase gene promoter. This can be used to generate RNA in a  
 CC tissue-specific or developmentally regulated manner, and this RNA  
 CC used to inhibit ACCase expression. Monocotyledonous plants can be made  
 CC resistant to grass-weed herbicides such as fluzafop and alkylketones  
 CC by transforming with cassettes adapted to express ACCase. This is  
 CC achieved by overexpression of monocot ACCase, expression of dicot  
 CC ACCase (which is relatively tolerant to these herbicides), or  
 CC possibly by expression of a resistant form of maize ACCase. This  
 CC provides crop plants which are resistant to herbicides used to  
 CC destroy weeds growing between the plants. This allows the overall  
 CC applicn. of herbicide without effecting the crop, and is also useful  
 CC where there has been short term carryover of herbicide from the  
 CC previous crop. Using the method, weeds such as wild oats may be  
 CC controlled, and the farmer is given extra options to obtain an improved  
 CC harvest using means which are safer, cheaper or more effective.  
 CC Suitable crops are maize, wheat and barley.  
 XX  
 SQ Sequence 4346 BP; 1239 A; 829 C; 1064 G; 1213 T; 1 other;

Query Match 73.6%; Score 16.2; DB 14; Length 4346;  
 Best Local Similarity 85.7%; Pred. No. 55;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 gttggatcacaaagtattgata 22  
 ||||| ||||| ||| ||

Db 902 gttggatgacaagagttgtta 922  
 RESULT 13  
 T39905  
 ID T39905 standard; cDNA; 7470 BP.  
 XX  
 XX T39905;  
 AC  
 XX 23-JAN-1997 (first entry)  
 DT  
 XX  
 XX Maize acetyl CoA carboxylase cDNA.  
 DE  
 XX  
 XX Acetyl CoA carboxylase; ACCase; herbicide tolerance;  
 KW cyclohexanedione; aryloxyphenoxypropionic acid; vegetable oil;  
 KW oilseed; maize; corn; ss.  
 KW  
 XX Zea mays line A188.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 37..7014  
 FT /\*tag= a  
 FT /EC\_number= 6.4.1.2  
 XX  
 XX W09631609-A2.  
 PN  
 XX  
 XX 10-OCT-1996.  
 PD  
 XX  
 XX 04-APR-1996; 96WO-US04625.  
 PF  
 XX  
 XX 05-APR-1995; 95US-0417089.  
 PR  
 XX  
 XX (MINU ) UNIV MINNESOTA.  
 PA  
 XX Egli MA, Gengenbach BG, Gronwald JW, Lutz SM, Somers DA;  
 PI Wyse DL;  
 PI  
 XX WPI: 1996-465030/46.  
 DR P-PSDB; W05590.  
 XX  
 XX DNA encoding maize acetyl coenzyme A carboxylase gene - used for  
 PT prodn. of plants with herbicide tolerance or altered oil content  
 PT  
 XX Claim 2; Page 78-80; 131pp; English.  
 PS  
 XX A cDNA sequence (T39905) codes for maize acetyl CoA carboxylase  
 CC (ACCase) (W05590), an enzyme that plays a central role in fatty  
 CC acid biosynthesis and accumulation in plants and seeds. The  
 CC complete sequence was deduced from cDNA clone #18-5, obt'd. from  
 CC a lambda gt10 cDNA library of maize inbred A188 seedlings, and from  
 CC PCR amplifications (see also T39918-23). Genomic ACCase DNA  
 CC fragments have also been isolated (see also T39906-16). ACCase DNA  
 CC can be incorporated into a vector and used to increase the  
 CC herbicide tolerance or oil content of a transgenic plant, or used  
 CC to produce ACCase in heterologous systems.  
 XX  
 XX Sequence 7470 BP; 2119 A; 1429 C; 1844 G; 2071 T; 7 other;  
 SQ  
 Query Match 73.6%; Score 16.2; DB 17; Length 7470;  
 Best Local Similarity 85.7%; Pred. No. 58;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 gttggatcacaaagtattgata 22  
 ||||| ||||| ||| ||  
 Db 3972 gttggatgacaagagttgtta 3992  
 RESULT 14  
 V29317  
 ID V29317 standard; cDNA; 7470 BP.  
 XX  
 XX V29317;

```
XX 30-JUL-1998 (first entry)
XX Maize ACCase enzyme encoding cDNA.
DE ACCase; maize; herbicide resistant; corn plant; tolerance: Acc1; Acc2;
KW cyclohexanedione; CHD; aryloxyphenoxypyranoic acid; APA; enzyme; ss.
XX Zea mays.
FH Key Location/Qualifiers
FT 37..7014
FT /*tag= a
FT /transl_except= (pos:4339..4341, aa: His)
FT /transl_except= (pos:4456..4458, aa: Ser)
FT /transl_except= (pos:4359..4461, aa: Lys)
FT /product= "ACCase"
XX W09808963-A1.
XX 05-MAR-1998.
XX 29-AUG-1997; 97WO-US15344.
XX 30-AUG-1996; 96US-0697826.
XX (MINU ) UNIV MINNESOTA.
XX Egli MA, Gengenbach BG, Lutz SM, Marshall LC, Parker WB;
PI Somers DA, Vandee KL, Wyse DL;
XX WPI; 1998-207043/18.
XX P-PSDB; W56736.
XX Herbicide resistant corn plants - prepared using Acc1 and Acc2 gene
XX combinations
XX Example 7; Fig 3A-E; 112pp; English.
XX This cDNA encodes a maize ACCase enzyme. This can be used in a method
XX of preparing an herbicide resistant corn plant which comprises crossing a
XX first corn plant to a second corn plant so as to yield a progeny plant,
XX where the first corn plant comprises at least 1 herbicide resistant
XX allele and the second plant comprises at least 1 herbicide resistant
XX allele which is not allelic to the herbicide resistant allele present in
XX the first plant. The herbicide resistant corn plants are prepared using
XX Acc1 and Acc2 gene combinations which impart cyclohexanedione (CHD) or
XX aryloxyphenoxypyranoic acid (APA) herbicide tolerance to the corn plant.
XX The methods are used to impart CHD and APA herbicide tolerance to corn
XX plants and to produce CHD or APA herbicide tolerant Zea mays (corn)
XX homozygous or heterozygous for Acc1 and homozygous or heterozygous for
XX Acc2. The methods can also be used to impart tolerance to a corn plant
XX to an agent which inhibits acetyl CoA carboxylase, selected from
XX 3-(2,4-dichlorophenyl)-perhydroindolizine-2,4-dione, 3-isopropyl-6-
XX (N-[2,2-dimethylpropyl]-acetamido-1,3,5-triazine-2,4 (1H,3H)dione,
XX soraphen A and their structural analogues.
XX Sequence 7470 BP; 2119 A; 1429 C; 1844 G; 2071 T; 7 other;

Query Match 73.6%; Score 16.2; DB 19; Length 7470;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggatcacaaagatttgata 22
DB 3972 gttggatgacaaagatttgata 3992
||||| ||||| |||||

RESULT 15
249820
ID 249820 standard; cDNA; 7470 BP.
XX
```

```
AC 249820;
XX 18-APR-2000 (first entry)
XX Maize acetyl CoA carboxylase cDNA.
DE Herbicide resistance; gene modification;
XX maize acetyl CoA carboxylase; ACCase; herbicide; plant oil content;
KW marker-assisted plant selection; groat oil trait;
KW restriction fragment length polymorphism mapping;
KW high-energy animal feed; low-fat human food; ss.
XX Zea mays.
FH Key Location/Qualifiers
FT 37..7014
FT /*tag= a
FT /product= "Maize acetyl CoA carboxylase"
FT 1..36
FT /*tag= b
FT 7014..7470
FT /*tag= c
XX W09967367-A1.
XX 29-DEC-1999.
XX 22-JUN-1999; 99WO-US14022.
XX 22-JUN-1998; 98US-0090240.
XX 02-JUL-1998; 98US-0091640.
XX (MINU ) UNIV MINNESOTA.
XX (USDA ) US DEPT OF AGRICULTURE.
XX (EGLI/) EGLI M A.
XX (GROH/) GROH S.
XX (KIAN/) KIANIAN S F.
XX (PHIL/) PHILLIPS R L.
XX (RINE/) RINES H W.
XX (SOME/) SOMERS D A.
XX Egli MA, Groh S, Kianian SF, Phillips RL, Rines HW, Somers DA;
PI WPI; 2000-147205/13.
XX P-PSDB; Y44687.
XX New DNA encoding acetyl-CoA carboxylase from oats, used to produce
XX transformed plants with herbicide resistance and altered oil content
XX Example 6; Fig 13; 197pp; English.
XX The present cDNA sequence encodes maize acetyl CoA carboxylase (ACCase).
XX ACCase is involved in fatty acid synthesis and is the target of
XX the specified herbicides. Transformation of plants with ACCase imparts
XX resistance to cyclohexanedione and aryloxyphenoxypyranoic acid
XX herbicides and alter the oil content. The ACCase DNA is also used as
XX source of probes and primers for the identification of transgenic
XX plants; in marker-assisted plant selection and for restriction fragment
XX length polymorphism mapping, used for high-energy animal feed and
XX high-fiber, low-fat human food and in genetic dissection of the groat
XX oil trait.
XX Sequence 7470 BP; 2119 A; 1430 C; 1843 G; 2071 T; 7 other;

Query Match 73.6%; Score 16.2; DB 21; Length 7470;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gttggatcacaaagatttgata 22
DB 3972 gttggatgacaaagatttgata 3992
||||| ||||| |||||
```

•  
Search completed: March 27, 2001, 08:23:02  
Job time: 5304 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 08:19:23 ; Search time 132.3 Seconds  
(without alignments)  
26.799 Million cell updates/sec

Title: US-09-380-826A-7  
Perfect score: 22  
Sequence: 1 tgttgatcacaaagtattgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/3B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.2	73.6	2000	1	US-08-014-326-1
2	16.2	73.6	2001	3	US-08-417-089-1
3	16.2	73.6	2001	3	US-08-695-651-1
4	16.2	73.6	4345	2	US-08-244-537-1
5	16.2	73.6	7470	3	US-08-417-089-5
6	16.2	73.6	7470	3	US-08-695-651-5
7	15.6	70.9	296	3	US-08-602-145-14
8	15.6	70.9	1096	1	US-08-684-862-8
9	15.6	70.9	1989	2	US-08-792-055-1
10	15.6	70.9	7745	1	US-08-299-675-1
11	15.6	70.9	7745	1	US-08-485-241-1
12	15.6	70.9	7745	2	US-08-874-162-1
13	15.2	69.1	1016	1	US-08-399-9868-3
14	15.2	69.1	1016	1	US-08-493-754A-3
15	15.2	69.1	2182	1	US-08-399-9868-1
16	15.2	69.1	2182	1	US-08-493-754A-1
17	15.2	69.1	2598	3	US-08-745-892-20
18	15.2	69.1	5420	5	5256642-3
19	15.2	69.1	5420	5	5472939-3
20	15.2	69.1	6951	5	5256642-1
21	15.2	69.1	6951	5	5472939-1
22	14.8	67.3	1855	3	US-08-961-083-71
23	14.8	67.3	3159	1	US-08-119-361-4
24	14.8	67.3	3159	3	US-08-336-308A-3
25	14.8	67.3	3159	3	US-08-822-324-3
26	14.8	67.3	7266	3	US-08-336-308A-9
27	14.8	67.3	7266	3	US-08-822-324-5
28	14.8	67.3	8640	1	US-08-570-311-28

29	14.8	67.3	9919	3	US-08-880-179-1	Sequence 1, Appli
30	14.6	66.4	293	3	US-08-866-340-13	Sequence 13, Appli
31	14.6	66.4	1404	1	US-08-204-656B-1	Sequence 1, Appli
32	14.6	66.4	1404	1	US-08-204-656B-3	Sequence 3, Appli
33	14.6	66.4	1404	1	US-08-204-656B-5	Sequence 5, Appli
34	14.6	66.4	1404	1	US-08-204-656B-7	Sequence 7, Appli
35	14.6	66.4	1404	1	US-08-470-702-1	Sequence 1, Appli
36	14.6	66.4	1404	1	US-08-470-702-2	Sequence 2, Appli
37	14.6	66.4	1404	1	US-08-470-702-3	Sequence 3, Appli
38	14.6	66.4	1404	1	US-08-470-702-4	Sequence 4, Appli
39	14.6	66.4	1404	1	US-08-467-831-1	Sequence 1, Appli
40	14.6	66.4	1404	1	US-08-467-831-2	Sequence 2, Appli
41	14.6	66.4	1404	1	US-08-467-831-3	Sequence 3, Appli
42	14.6	66.4	1404	1	US-08-467-831-4	Sequence 4, Appli
43	14.6	66.4	1608	2	US-08-424-224-1	Sequence 1, Appli
44	14.6	66.4	1608	4	PCT-US94-02891-68	Sequence 68, Appli
45	14.6	66.4	12047	2	US-09-022-461-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-014-326-1  
; Sequence 1, Application US/08014326  
; Patent No. 5498544  
; GENERAL INFORMATION:  
; APPLICANT: Gengenbach, Burle G.  
; APPLICANT: Somers, David A.  
; APPLICANT: Wyse, Donald L.  
; APPLICANT: Gronwald, John W.  
; APPLICANT: Egli, Margaret A.  
; APPLICANT: Lutz, Shleia M.  
; TITLE OF INVENTION: Method and An Acetyl CoA Carboxylase  
; TITLE OF INVENTION: Alteration in Oil Content of Plants  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 3100 No. 5498544west Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/014,326  
FILING DATE: 05-FEB-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,462  
FILING DATE: 21-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/538,674  
FILING DATE: 18-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/269,584  
FILING DATE: 10-NOV-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Kowalchuk, Katherine M.  
REGISTRATION NUMBER: 36,848  
REFERENCE/DOCKET NUMBER: 600.258-US01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2000 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

```
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: 2 kb fragment of lambda clone #15-14
US-08-014-326-1

Query Match 73.6%; Score 16.2; DB 1; Length 2000;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagattgata 22
   ||||| ||||| ||| ||
Db 42 GTTGGATGACAAGAGTTGTTA 62

RESULT 2
US-08-417-089-1
; Sequence 1, Application US/08417089
; Patent No. 6069298
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN
; TITLE OF INVENTION: OIL CONTENT OF PLANTS
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-417-089-1

Query Match 73.6%; Score 16.2; DB 3; Length 2001;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagattgata 22
   ||||| ||||| ||| ||
Db 42 GTTGGATGACAAGAGTTGTTA 62

RESULT 3
US-08-695-651-1
; Sequence 1, Application US/08695651
; Patent No. 6146867
; GENERAL INFORMATION:
; APPLICANT: Gengenbach, B. G.
; APPLICANT: Somers, D. A.
; APPLICANT: Wyse, D. L.
; APPLICANT: Gronwald, J. W.
; APPLICANT: Egli, M. A.
; APPLICANT: Lutz, S. M.
; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schweeman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/695,651
; FILING DATE: 12-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/417089
; FILING DATE: 05-APR-1995
; APPLICATION NUMBER: 08/014326
; FILING DATE: 05-FEB-1993
; APPLICATION NUMBER: 07/917462
; FILING DATE: 21-JUL-1992
; APPLICATION NUMBER: 07/538674
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.318US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-373-6900
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-695-651-1

Query Match 73.6%; Score 16.2; DB 3; Length 2001;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacaaagattgata 22
   ||||| ||||| ||| ||
Db 42 GTTGGATGACAAGAGTTGTTA 62

RESULT 4
US-08-244-537-1
; Sequence 1, Application US/08244537
; Patent No. 5854420
; GENERAL INFORMATION:
; APPLICANT: ASHTON, ANTHONY R.
; APPLICANT: JENKINS, COLIN L.D.
; APPLICANT: WHITFIELD, PAUL R.
; TITLE OF INVENTION: MAIZE ACETYL COA CARBOXYLASE ENCODING
; TITLE OF INVENTION: DNA CLONES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,537
; FILING DATE: 18-AUG-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9125330.2
```



;; FILING DATE: 28-NOV-1991  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/GB92/02205  
;; FILING DATE: 27-NOV-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: KOKULIS, PAUL N.  
;; REGISTRATION NUMBER: 16,773  
;; REFERENCE/DOCKET NUMBER: 203094/SEE 36663/UST  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 861-3000  
;; TELEFAX: (202) 822-0944  
;; TELEX: 6714627 CUSH  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4345 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
US-08-244-537-1

Query Match 73.6%; Score 16.2; DB 2; Length 4345;  
Best Local Similarity 85.7%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22  
||||| ||||| ||| ||  
Db 902 GTTGGATGACAGAGTTGTTA 922

RESULT 5  
US-08-417-089-5  
; Sequence 5, Application US/08417089  
; Patent No. 6059298  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE  
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN  
; TITLE OF INVENTION: OIL CONTENT OF PLANTS  
; NUMBER OF SEQUENCES: 11  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/417,089  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7470 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-417-089-5

Query Match 73.6%; Score 16.2; DB 3; Length 7470;  
Best Local Similarity 85.7%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22  
||||| ||||| ||||| ||| ||  
Db 3972 GTTGGATGACAGAGTTGTTA 3992

RESULT 6  
US-08-695-651-5  
; Sequence 5, Application US/08695651  
; Patent No. 6146867  
; GENERAL INFORMATION:  
; APPLICANT: Gengenbach, B. G.

;; APPLICANT: Somers, D. A.  
;; APPLICANT: Wyse, D. L.  
;; APPLICANT: Gronwald, J. W.  
;; APPLICANT: Egli, M. A.  
;; APPLICANT: Latz, S. M.  
;; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA  
;; NUMBER OF SEQUENCES: 11  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.  
;; STREET: P.O. Box 2938  
;; CITY: Minneapolis  
;; STATE: MN  
;; COUNTRY: USA  
;; ZIP: 55402  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSEQ Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/695,651  
;; FILING DATE: 12-AUG-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/417089  
;; FILING DATE: 05-APR-1995  
;; APPLICATION NUMBER: 08/014326  
;; FILING DATE: 05-FEB-1993  
;; APPLICATION NUMBER: 07/917462  
;; FILING DATE: 21-JUL-1992  
;; APPLICATION NUMBER: 07/538674  
;; FILING DATE: 18-JUN-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Woessner, Warren D  
;; REGISTRATION NUMBER: 30,440  
;; REFERENCE/DOCKET NUMBER: 600.318US3  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 612-373-6900  
;; TELEFAX: 612-339-3061  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 7470 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
US-08-695-651-5

Query Match 73.6%; Score 16.2; DB 3; Length 7470;  
Best Local Similarity 85.7%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gttggatcacagaatttgata 22  
||||| ||||| ||||| ||| ||  
Db 3972 GTTGGATGACAGAGTTGTTA 3992

RESULT 7  
US-08-602-145-14/c  
; Sequence 14, Application US/08602145  
; Patent No. 6025336  
; GENERAL INFORMATION:  
; APPLICANT: Goltry, Kristin L.  
; APPLICANT: Greenberger, Joel S.  
; TITLE OF INVENTION: DETERMINING EXPOSURE TO IONIZING RADIATION  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington

```

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Aekistrodon rhodostoma
; LOCATION: 144 to 841
; OTHER INFORMATION: the coding region shown in (2)(ix)(B)
; OTHER INFORMATION: codes for the protein of SEQ ID NO: 3
;
; US-08-684-862-8
;
; Query Match 70.9%; Score 15.6; DB 1; Length 1096;
; Best Local Similarity 81.8%; Pred. No. 37;
; Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 tggtagatcacagatttgata 22
; ||| | ||||| |||| ||
;
; DB 112 TGTAGTATCACAGGTTTGCTA 91
;
;
; RESULT 9
; US-08-792-055-1/c
; ; Sequence 1, Application US/08792055
; ; Patent No. 5853980
; ; GENERAL INFORMATION:
; ; APPLICANT: Rollin, Pierre E.
; ; APPLICANT: Elliott, Luanne
; ; APPLICANT: Ksiazek, Thomas G.
; ; APPLICANT: Nichol, Stuart T.
; ; APPLICANT: Morzunov, Sergey
; ; APPLICANT: Ravkov, Eugene
; ; TITLE OF INVENTION: The Black Creek Canal Hantavirus and
; ; TITLE OF INVENTION: Related Methods
; ; NUMBER OF SEQUENCES: 7
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; ; STREET: 127 Peachtree Street, N.E., Suite 1200
; ; CITY: Atlanta
; ; STATE: Georgia
; ; COUNTRY: USA
; ; ZIP: 30303
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: PatentIn Release #1.0, Version #1.25
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/08/792.055
; ; FILING DATE: 03-FEB-1997
; ; CLASSIFICATION: 435
; ; PRIOR APPLICATION DATA:
; ; APPLICATION NUMBER: 08/390,361
; ; FILING DATE:
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Spratt, Gwendolyn D.
; ; REGISTRATION NUMBER: 36,016
; ; REFERENCE/DOCKET NUMBER: 1414.622
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (404) 688-0770
; ; TELEFAX: (404) 688-9880
; ; INFORMATION FOR SEQ ID NO: 1:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 1989 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: double
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: DNA (genomic)
; ; US-08-792-055-1

```

Query Match 70.9%; Score 15.6; DB 2; Length 1989;  
Best Local Similarity 81.8%; Pred. No. 39;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagatttgata 22  
||||| ||||| ||  
Db 288 TGTGGATCAACAGATTGTA 267

RESULT 10  
US-08-299-675-1/c  
; Sequence 1, Application US/08299675  
; Patent No. 5589618  
; GENERAL INFORMATION:  
; APPLICANT: Hannah, L. Curtis  
; APPLICANT: Giroux, Michael  
; TITLE OF INVENTION: Materials and Methods for Increasing  
; TITLE OF INVENTION: Corn Seed Weight  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/299,675  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: UF146  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7745 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-299-675-1

Query Match 70.9%; Score 15.6; DB 1; Length 7745;  
Best Local Similarity 81.8%; Pred. No. 46;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagatttgata 22  
||||| ||||| ||  
Db 3600 TGTGGATCAACAGATGACATA 3579

RESULT 11  
US-08-485-241-1/c  
; Sequence 1, Application US/08485241  
; Patent No. 5650557  
; GENERAL INFORMATION:  
; APPLICANT: Hannah, L. Curtis  
; APPLICANT: Giroux, Michael  
; TITLE OF INVENTION: Materials and Methods for Increasing  
; TITLE OF INVENTION: Corn Seed Weight  
; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,241  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/299,675  
; FILING DATE: 1-SEP-1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pace, Doran R.  
; REGISTRATION NUMBER: 38,261  
; REFERENCE/DOCKET NUMBER: UF146.C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7745 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-485-241-1

Query Match 70.9%; Score 15.6; DB 1; Length 7745;  
Best Local Similarity 81.8%; Pred. No. 46;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagatttgata 22  
||||| ||||| ||  
Db 3600 TGTGGATCAACAGATGACATA 3579

RESULT 12  
US-08-874-162-1/c  
; Sequence 1, Application US/08874162  
; Patent No. 5872216  
; GENERAL INFORMATION:  
; APPLICANT: Hannah, L. Curtis  
; APPLICANT: Giroux, Michael  
; TITLE OF INVENTION: Materials and Methods for Increasing  
; TITLE OF INVENTION: Corn Seed Weight  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/874,162  
; FILING DATE: 13-JUN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:

```

* APPLICATION NUMBER: US 08/485,241
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/299,675
; FILING DATE: 1-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Doran R.
; REGISTRATION NUMBER: 38,261
; REFERENCE/DOCKET NUMBER: UF-146C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7745 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-874-162-1

Query Match 70.9%; Score 15.6; DB 2; Length 7745;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttgtgatcacaaagtattgata 22
||||| ||||| ||||| |||||
Db 3600 TGTGGATAACAAGATGACATA 3579

RESULT 13
US-08-399-986B-3/c
; Sequence 3, Application US/08399986B
; Patent No. 5801041
; GENERAL INFORMATION:
; APPLICANT: Godwin, Andrew K.
; TITLE OF INVENTION: No. 5801041el Gene Associated with Suppression
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,986B
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

```

US-08-399-986B-3

Query Match 69.1%; Score 15.2; DB 1; Length 1016;
Best Local Similarity 85.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagtattgata 22
||||| ||||| ||||| |||||
Db 830 TTGGGTCTCAAGAATTGATA 811

RESULT 14
US-08-493-754A-3/c
; Sequence 3, Application US/08493754A
; Patent No. 5821338
; GENERAL INFORMATION:
; APPLICANT: Godwin, Andrew K.
; TITLE OF INVENTION: No. 5821338el Gene Associated with Suppression
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/493,754A
; FILING DATE: 22-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-493-754A-3

Query Match 69.1%; Score 15.2; DB 1; Length 1016;
Best Local Similarity 85.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttggatcacaaagtattgata 22
||||| ||||| ||||| |||||
Db 830 TTGGGTCTCAAGAATTGATA 811

RESULT 15
US-08-399-986B-1/c
; Sequence 1, Application US/08399986B
; Patent No. 5801041
; GENERAL INFORMATION:
; APPLICANT: Godwin, Andrew K.
; TITLE OF INVENTION: No. 5801041el Gene Associated with Suppression
; NUMBER OF SEQUENCES: 35

```

```

; - CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,986B
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2182 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-399-986B-1

```

```

Query Match      69.1%; Score 15.2; DB 1; Length 2182;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      3 ttggatcacaaagatttgata 22
        |||| || |||| |||||
DB      1996 TTGGGTCTCAGAATTGATA 1977

```

```

Search completed: March 27, 2001, 08:19:25
Job time: 5152 sec

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2001, 07:38:25 ; Search time 2517.78 Seconds  
(without alignments)  
61.230 Million cell updates/sec

Title: US-09-380-826A-7  
Perfect score: 22  
Sequence: 1 tggatgacacaaagatttgata 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estp11:\*  
70: em\_estp12:\*  
71: em\_estp13:\*  
72: em\_estp14:\*  
73: em\_estp15:\*  
74: em\_estrol1:\*  
75: em\_estrol2:\*  
76: em\_estrol3:\*  
77: em\_estrol4:\*  
78: em\_estrol5:\*  
79: em\_estrol6:\*  
80: em\_estrol7:\*  
81: em\_estrol8:\*  
82: em\_estrol9:\*  
83: em\_estrol10:\*  
84: em\_estrol11:\*  
85: em\_estrol12:\*  
86: em\_estrol13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estpl6:\*  
118: em\_estpl7:\*  
119: em\_estpl8:\*  
120: em\_estpl9:\*  
121: em\_estpl10:\*  
122: em\_estpl11:\*  
123: em\_estpl12:\*  
124: em\_estpl13:\*  
125: em\_estpl14:\*  
126: em\_estpl15:\*  
127: em\_estpl16:\*  
128: em\_estpl17:\*  
129: em\_estpl18:\*  
130: em\_estpl19:\*  
131: em\_estpl20:\*  
132: em\_estpl21:\*  
133: em\_estpl22:\*  
134: em\_estpl23:\*  
135: em\_estpl24:\*  
136: em\_estpl25:\*  
137: em\_estpl26:\*  
138: em\_estpl27:\*  
139: em\_estpl28:\*  
140: em\_estpl29:\*  
141: em\_estpl30:\*  
142: em\_estpl31:\*  
143: em\_estpl32:\*  
144: em\_estpl33:\*  
145: em\_estpl34:\*  
146: em\_estpl35:\*  
147: em\_estpl36:\*  
148: em\_estpl37:\*  
149: em\_estpl38:\*  
150: em\_estpl39:\*  
151: em\_estpl40:\*  
152: em\_estpl41:\*  
153: em\_estpl42:\*  
154: em\_estpl43:\*  
155: em\_estpl44:\*  
156: em\_estpl45:\*  
157: em\_estpl46:\*  
158: em\_estpl47:\*  
159: em\_estpl48:\*  
160: em\_estpl49:\*  
161: em\_estpl50:\*  
162: em\_estpl51:\*  
163: em\_estpl52:\*  
164: em\_estpl53:\*  
165: em\_estpl54:\*  
166: em\_estpl55:\*  
167: em\_estpl56:\*  
168: em\_estpl57:\*  
169: em\_estpl58:\*  
170: em\_estpl59:\*  
171: em\_estpl60:\*  
172: em\_estpl61:\*  
173: em\_estpl62:\*  
174: em\_estpl63:\*  
175: em\_estpl64:\*  
176: em\_estpl65:\*  
177: em\_estpl66:\*  
178: em\_estpl67:\*  
179: em\_estpl68:\*  
180: em\_estpl69:\*  
181: em\_estpl70:\*  
182: em\_estpl71:\*  
183: em\_estpl72:\*  
184: em\_estpl73:\*  
185: em\_estpl74:\*  
186: em\_estpl75:\*  
187: em\_estpl76:\*  
188: em\_estpl77:\*  
189: em\_estpl78:\*

190: gb\_gss25:\*  
191: gb\_gss26:\*  
192: gb\_gss27:\*  
193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.8	85.5	361	37	AV532040	AV532040 AV532040
C 2	18.8	85.5	394	147	226559	226559 ATTS1651 Ve
C 3	18.8	85.5	490	14	AB038725	AB038725 AB038725
C 4	18.8	85.5	527	37	AV520789	AV520789 AV520789
C 5	18.8	85.5	530	27	AI995637	AI995637 701676626
C 6	18.8	85.5	608	138	BE977618	BE977618 bs66808.y
C 7	18.4	83.6	934	191	CNS02BWA	AL190387 Tetraodon
C 8	17.8	80.9	633	138	BF006254	BF006254 EST434942
C 9	17.8	80.9	633	138	BF006254	BF006254 EST434942
C 10	17.4	79.1	286	126	BB288727	BB288727 BB288727
C 11	17.4	79.1	407	110	BE581720	BE581720 kg51c07.y
C 12	17.4	79.1	529	158	AQ437779	AQ437779 HS_5066.B
C 13	17.4	79.1	674	94	AW695022	AW695022 NF082E04S
C 14	17.4	79.1	762	106	BE283043	BE283043 601101323
C 15	17.4	79.1	837	110	BE642809	BE642809 Cr12.7_B0
C 16	17.4	79.1	1101	190	CNS00D26	AL077427 Drosophila
C 17	17.2	78.2	151	134	BE068225	BE068225 MR4-BF036
C 18	17.2	78.2	425	15	AI049911	AI049911 an30h03.x
C 19	17.2	78.2	431	159	AQ595814	AQ595814 HS_2132.B
C 20	17.2	78.2	432	150	AQ221715	AQ221715 HS_2010.A
C 21	17.2	78.2	437	146	W43212	W43212 22591_Lambd
C 22	17.2	78.2	439	162	AQ796989	AQ796989 nbxb0071H
C 23	17.2	78.2	500	24	AI733664	AI733664 an30h03.x
C 24	17.2	78.2	524	174	B62582	B62582 T22F18TR.TA
C 25	17.2	78.2	531	150	AQ234477	AQ234477 HS_2003.B
C 26	17.2	78.2	536	39	AW034253	AW034253 EST277824
C 27	17.2	78.2	546	24	AI779714	AI779714 EST260593
C 28	17.2	78.2	579	174	B67515	B67515 T22M10TR.TA
C 29	17.2	78.2	602	24	AI777095	AI777095 EST258060
C 30	17.2	78.2	614	105	BE187570	BE187570 EST336131
C 31	17.2	78.2	619	97	AW963686	AW963686 EST375759
C 32	17.2	78.2	680	174	B57784	B57784 CIT-HSP-201
C 33	17.2	78.2	698	168	A2193715	A2193715 SP_1023.B
C 34	17.2	78.2	904	191	CNS029DN	AL187124 Tetraodon
C 35	17.2	78.2	949	192	CNS04HNP	AL291166 Tetraodon
C 36	17.3	77.3	373	140	C70416	C70416 C70416.YUJ1
C 37	16.8	76.4	238	34	AV313332	AV313332 AV313332
C 38	16.8	76.4	260	104	BE118999	BE118999 UI-R-CA0-
C 39	16.8	76.4	296	145	T20362	T20362 6c01908-t7
C 40	16.8	76.4	396	97	AW943911	AW943911 LD47517.3
C 41	16.8	76.4	432	157	AQ449764	AW487226 B1727.MAR
C 42	16.8	76.4	446	24	AI761541	AQ449764 500004C01
C 43	16.8	76.4	446	24	AI761541	AI761541 w161f07.x
C 44	16.8	76.4	450	138	BF002933	BF002933 7g51c09.x
C 45	16.8	76.4	464	156	AQ332265	AQ332265 HS_5007.A

ALIGNMENTS

RESULT 1  
AV532040/c  
LOCUS AV532040 361 bp mRNA  
DEFINITION Arabidopsis thaliana flower buds Columbia Arabidopsis thaliana cDNA clone FB034f09f 3', mRNA sequence.  
ACCESSION AV532040  
VERSION AV532040.1 GI:8692323  
KEYWORDS EST  
SOURCE thale cress.



```

ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE      1 (bases 1 to 361)
AUTHORS        Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE          A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL        DNA Res. 7, 175-180 (2000)
MEDLINE        20363093
COMMENT        Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..361
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="FB034f09f"
/clone_lib="Arabidopsis thaliana flower buds Columbia"
/tissue_type="flower buds"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT    113 a 73 c 61 g 114 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 37; Length 361;
Best Local Similarity 90.9%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ttgtggtcaccaagatttgata 22
||||| |||||||||
Db      276 TGTGTGTTCCACAGATTTCACA 255

RESULT 2
LOCUS      226559 394 bp mRNA EST 31-MAY-1995
DEFINITION ATTS1651 Versailles-VC Arabidopsis thaliana cDNA clone VCVDH08 3'
similar to Ribonuclease (RNS2), mRNA sequence.
ACCESSION 226559
VERSION 226559.1 GI:404227
KEYWORDS EST.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE    1 (bases 1 to 394)
AUTHORS      CNRS.
TITLE        The Arabidopsis thaliana transcribed genome: the GDR cDNA program
Unpublished (1996)
JOURNAL      Contact: Desprez F., Anselm J., Chiapello H., Rouze P., Caboche
M., Hofte H.
COMMENT      INRA Versailles
Laboratoire de Biologie Cellulaire
Route de Saint-Cyr, 78026 Versailles Cedex, France
Email: thierry@versailles.inra.fr.
Location/Qualifiers
1..394
/organism="Arabidopsis thaliana"
/strain="ecotype Columbia"
/db_xref="taxon:3702"
/clone_lib="VCVDH08"
/clone_lib="Versailles-VC"
/tissue_type="whole seedlings"
/dev_stage="in vitro-grown etiolated seedlings, 5 days old"
/note="Vector: pBluescript"
BASE COUNT   117 a 80 c 63 g 133 t
1 others

```

```

ORIGIN
Query Match      85.5%; Score 18.8; DB 147; Length 394;
Best Local Similarity 90.9%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ttgtggtcaccaagatttgata 22
||||| |||||||||
Db      321 TGTGTGTTCCACAGATTTCACA 300

RESULT 3
LOCUS      AB038725 490 bp mRNA EST 29-SEP-2000
DEFINITION AB038725 Arabidopsis thaliana Above-ground organ from two to
six-week old plants Columbia Arabidopsis thaliana cDNA clone
AP230g03_f 3', mRNA sequence.
ACCESSION  AB038725
VERSION    AB038725
KEYWORDS   EST.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE    1 (bases 1 to 490)
AUTHORS      Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE        A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL      DNA Res. 7, 175-180 (2000)
MEDLINE      20363093
COMMENT      Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..490
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="AP230g03_f"
/clone_lib="Arabidopsis thaliana Above-ground organ from
two to six-week old plants Columbia"
/tissue_type="Above-ground organ from two to six-week old
plants"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT    147 a 99 c 87 g 157 t
ORIGIN
Query Match      85.5%; Score 18.8; DB 14; Length 490;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ttgtggtcaccaagatttgata 22
||||| |||||||||
Db      303 TGTGTGTTCCACAGATTTCACA 282

RESULT 4
LOCUS      AV520789/c 527 bp mRNA EST 01-SEP-2000
DEFINITION AV520789 Arabidopsis thaliana aboveground organs two to six-week
old Arabidopsis thaliana cDNA clone AP232g06f 3', mRNA sequence.
ACCESSION  AV520789
VERSION    AV520789.1 GI:8680316
KEYWORDS   EST.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana

```

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

## REFERENCE

## AUTHORS

1 (bases 1 to 527)

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

## JOURNAL

## MEDLINE

20363093

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

source

1..527

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="AP232g06F"

/clone\_lib="Arabidopsis thaliana aboveground organs two to six-week old"

/tissue\_type="aboveground organs"

/dev\_stage="two to six-week old"

/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

## BASE COUNT

157 a 111 c 101 g 158 t

## ORIGIN

Query Match 85.5%; Score 18.8; DB 37; Length 527;  
Best Local Similarity 90.9%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagtattgata 22

||||| |||||||||

Db 247 TGTGGTTCACAAAGATTGACA 226

## RESULT 5

AI995637/c

LOCUS

701676626 A. thaliana, Columbia Col-0, inflorescence-1 Arabidopsis thaliana cDNA clone 701676626, mRNA sequence.

AI995637

AI995637.1 GI:5842542

EST

thale cress.

ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

## REFERENCE

## AUTHORS

1 (bases 1 to 530)  
Chen, J., Momiyama, M., Chan, E., Mooney, M., Cartoon, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobruga, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.  
Arabidopsis thaliana Gene Expression Microarray  
Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: David Smoller, Ph.D.

Genome Systems, Inc., a wholly owned subsidiary of Incyte

Pharmaceuticals, Inc.

4633 World Parkway Circle, St. Louis, MO 63134, USA

Tel: 877-577-2733

Fax: 314-427-3324

Email: service@genomesystems.com.

Location/Qualifiers

1..530

/organism="Arabidopsis thaliana"

/cultivar="Columbia Col-0"

## FEATURES

source

/db\_xref="taxon:3702"  
/clone="701676626"  
/clone\_lib="A. thaliana, Columbia Col-0, inflorescence-1"  
/tissue\_type="inflorescence"  
/dev\_stage="4 - 7 weeks"  
/note="Vector: pSPORT; Site\_1: NotI; Site\_2: SalI; cDNA library was derived from untreated inflorescence tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunted, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."  
BASE COUNT: 163 a 108 c 100 g 158 t 1 others  
ORIGIN

Query Match 85.5%; Score 18.8; DB 27; Length 530;  
Best Local Similarity 90.9%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttttgatcacaaagtattgata 22

||||| |||||||||

Db 254 TGTGGTTCACAAAGATTGACA 233

## RESULT 6

BE977618/c

LOCUS

BE977618 608 bp mRNA EST 04-OCT-2000  
DEFINITION BS666H08.y1 Drosophila melanogaster adult testis library Drosophila melanogaster cDNA clone bs66H08 5', mRNA sequence.

BE977618

BE977618.1 GI:10608272

EST

fruit fly.

ORGANISM

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Brian Oliver

Laboratory of Cellular and Developmental Biology

NIDDK, National Institutes of Health

6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA

Fax: (301) 496 5239

Email: oliver@helix.nih.gov,

http://www.niddk.nih.gov/intram/people/boliver.htm

Tissue isolation and library construction performed at the National Institute of Diabetes and Digestive and Kidney Diseases, NIH (see http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC; see http://www.nisc.nih.gov).

Plate: 66 row: h column: 08

Seq primer: M13RP1 reverse primer (ABI).

FEATURES

source

1..608

/organism="Drosophila melanogaster"

/strain="y[+] w[67cl]/Y"

/db\_xref="taxon:7227"

/clone="bs66H08"

/clone\_lib="Drosophila melanogaster adult testis library"

/sex="male"

/dev\_stage="1-5 day adult"

/lab\_host="SOLR (Stratagene)"

/note="Organ: testis; Vector: pBlueScript SK (Stratagene); Site\_1: EcoR I; Site\_2: Xho I; Testes dissected from 1-5 day adult y[+] w[67cl]/Y males raised at 25oc. RNA isolated using Trizol (Life Technologies) and a single

round of Poly(A)+ selection using Oligotex (Qiagen). cDNA library constructed using Stratagene ZAP-cDNA synthesis kit. Oligo dt-primed, size fractionated ~1-6 kb, and directionally cloned at EcoRI and XhoI in Uni-ZAP XR. Following a single round of amplification pBluescript SK phagemids were mass excised. A distribution channel for clones is being sought, but not currently available. Requests for clones cannot be honored.

BASE COUNT 173 a 146 c 161 g 128 t

Query Match 83.6%; Score 18.4; DB 138; Length 608;  
Best Local Similarity 95.0%; Pred. No. 59;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ttggatcacaaagtattgata 22  
||||| |||||||

Db 74 TTGGATCCCAAGATTGATA 55

RESULT 7  
CNS02BWA/c

LOCUS  
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 254C22 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL190387

VERSION AL190387.1 GI:7828491

KEYWORDS GSS: genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha; Holacanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE

AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 934)  
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 934)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES

source

1..934

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="254C22"

/clone\_lib="G"

/note="Genoscope sequence ID : C0AG254BB1SP1-end ;

PUC-ori"

BASE COUNT 282 a 175 c 207 g 259 t 11 others

ORIGIN

Query Match 83.6%; Score 18.4; DB 191; Length 934;  
Best Local Similarity 95.0%; Pred. No. 62;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ttggatcacaaagtattgata 22  
||||| |||||||

Db 840 TTGGAACACAGATTGATA 821

RESULT 8  
BF006444/c

LOCUS  
DEFINITION EST434942 DSLC Medicago truncatula cDNA clone pDSLC-41J15, mRNA sequence.

ACCESSION BF006444

VERSION BF006444.1 GI:10706719

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE 1 (bases 1 to 600)

AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Gonzales,M.B. and Ellis,L.

TITLE ESTs from Medicago truncatula leaves and cotyledons

JOURNAL Unpublished (2000)

COMMENT Contact: Deborah A. Samac

Department of Plant Pathology

University of Minnesota

495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA

Tel: 612 625 1243

Fax: 651 649 5058

Email: debbys@pucini.crl.umn.edu

University of Minnesota name: M275482e TIGR sequence name:

MTLBA567K More information is available at:

<http://chrysis.tamu.edu/medicago>

Seq primer: SKmod (CTA gaa CTA gtg gat CC).

FEATURES

source

1..600

/organism="Medicago truncatula"

/cultivar="genotype A17"

/db\_xref="taxon:3880"

/clone="pDSLC-41J15"

/clone\_lib="DSLC"

/tissue\_type="leaves and cotyledons"

/dev\_stage="mixture of cotyledons from five days old plants and leaves obtained from two weeks old plants"

/lab\_host="E. coli strain SOLR"

/note="Vector: pBluescript SK +/-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of cotyledons of five days old plants and leaves of two weeks old plants. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 127 a 120 c 132 g 221 t

ORIGIN

Query Match 80.9%; Score 17.8; DB 138; Length 600;

Best Local Similarity 90.5%; Pred. No. 1.2e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tggatcacaaagtattgata 21  
||||| |||||||

Db 90 TTTTGATCACAGATTAGAT 70

RESULT 9

BF006254/c

LOCUS

DEFINITION EST434752 DSLC Medicago truncatula cDNA clone pDSLC-40G21, mRNA



RESULT 11	BE581720/c	16-AUG-2000
LOCUS	BE581720	EST
DEFINITION	kq51c07.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to wp:TL910.2 CE16413 ;, mRNA sequence.	
ACCESSION	BE581720	GI:98326562
VERSION	EST.	
KEYWORDS	Strongyloides stercoralis.	
SOURCE	Strongyloides stercoralis	
ORGANISM	Strongyloides stercoralis	
REFERENCE	1 (bases 1 to 407)	
AUTHORS	McCarter, J., Clifton, S., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.	
TITLE	The Washington Univ. Nematode EST Project, 1999	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: McCarter JP The Washington Univ. Nematode EST Project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu The library was constructed by Dr. Thomas Nutman and colleagues of NIAID, NIH (nutman@nih.gov). DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. High quality sequence stop: 384.	
FEATURES	source	
	Location/Qualifiers	
	1..407	
	/organism="Strongyloides stercoralis"	
	/strain="Rhabditiform larvae obtained from gerbils"	
	/db_xref="taxon:6248"	
	/clone_lib="TBN95TM-SSR"	
	/lab_host="XL-1 Blue MRP" (Stratagene)"	
	/note="vector: Lambda Uni-ZAP XR (Stratagene); Site_1: EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3 rhabditiform larvae which had been isolated from gerbils experimentally infected with larvae originally isolated from experimentally infected dogs. cDNA was constructed and, using adaptors, was cloned unidirectionally into the vector from the EcoRI site to the XhoI site. The library has an unamplified titer of 1 x 10E5 pfu/ml and an amplified, undiluted titer of 9 x 10E11 pfu/ml. The average insert size of the unamplified library is 675 bp (range, 100-1700)."	
BASE COUNT	152 a 63 c 74 g 118 t	
ORIGIN		
	Query Match 79.1%; Score 17.4; DB 110; Length 407;	
	Best Local Similarity 94.7%; Pred. No. 1.8e+02;	
	Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	4 tggatcacagaatttgata 22	
Db	334 TTGATCACAGATTGTGATA 316	
RESULT 12	AQ497779	28-APR-1999
LOCUS	AQ497779	GSS
DEFINITION	HS_5066_B1_E05_TTA RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=642 Col=9 Row=J, DNA sequence.	
ACCESSION	AQ497779	GI:4697902
VERSION	GSS.	
KEYWORDS	human.	
SOURCE	Homo sapiens	
ORGANISM	Homo sapiens	
REFERENCE	1 (bases 1 to 529)	
AUTHORS	Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.	
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	
MEDLINE	99380589	
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 642 row: J column: 9 Seg primer: T7 Class: BAC ends High quality sequence stop: 529.	
FEATURES	source	
	Location/Qualifiers	
	1..529	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone_lib="Plate=642 Col=9 Row=J"	
	/sex="male"	
	/note="vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"	
BASE COUNT	145 a 99 c 109 g 165 t	
ORIGIN		
	Query Match 79.1%; Score 17.4; DB 158; Length 529;	
	Best Local Similarity 94.7%; Pred. No. 1.8e+02;	
	Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	3 ttggatcacagaatttgat 21	
Db	403 TTGGGTCAACAGATTGTGAT 421	
RESULT 13	AW695022	15-JUN-2000
LOCUS	AW695022	EST
DEFINITION	NF082E04Srt1F1034 Developing stem Medicago truncatula cDNA clone NF082E04ST 5', mRNA sequence.	
ACCESSION	AW695022	
VERSION	EST.	
KEYWORDS	barrel medic.	
SOURCE	Medicago truncatula	
ORGANISM	Medicago truncatula	
REFERENCE	1 (bases 1 to 674)	
AUTHORS	He, X.-Z., Shadle, G., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Dixon, R.A.	
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula stem library	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Dixon RA	





**THIS PAGE BLANK (USPTO)**